

N O T I C E

THIS DOCUMENT HAS BEEN REPRODUCED FROM
MICROFICHE. ALTHOUGH IT IS RECOGNIZED THAT
CERTAIN PORTIONS ARE ILLEGIBLE, IT IS BEING RELEASED
IN THE INTEREST OF MAKING AVAILABLE AS MUCH
INFORMATION AS POSSIBLE

MITRE-Bedford

MTR - 8164

NASA CR 165155

(NASA-CR-165155) SECOND YEAR TECHNICAL
REPORT ON-BOARD PROCESSING FOR FUTURE
SATELLITE COMMUNICATIONS SYSTEMS (Mitre
Corp.) 255 p HC A10/MF A01

N81-10242

CSCL 17B

Unclass

G3/32 29083

Second Year Technical Report On-Board Processing for Future Satellite Communications Systems

William T. Brandon
Warren K. Green
Murray Hoffman
Paul N. Jean
William R. Neal
Brian E. White

October 1980



This document was prepared for authorized distribution.
It has not been approved for public release.

MITRE

MITRE Technical Report

MTR-8164

Second Year Technical Report On-Board Processing for Future Satellite Communications Systems

**William T. Brandon
Warren K. Green
Murray Hoffman
Paul N. Jean
William R. Neal
Brian E. White**

October 1980

CONTRACT SPONSOR
CONTRACT NO
PROJECT NO
DEPT.

NASA Lewis Research Center/NASA Report No. CR-165155
F19628-80-C-0001
8680
D-97



This document was prepared for authorized distribution.
It has not been approved for public release.

Department Approval: W.T. Brandon

MITRE Project Approval: W.T. Brandon

ABSTRACT

Advanced baseband and microwave switching techniques for large domestic communications satellites operating in the 30/20 GHz frequency bands are discussed. A baseline design employing the same 100 MHz bandwidth in each of 16 movable spot antenna beams resulted in a capacity of 1.6 Gb/s for a regenerative baseband processor power consumption of 1 kW. Link establishment and packet routing protocols are defined. Also described is a detailed design of a separate 100 x 100 microwave switch capable of handling non-regenerated signals occupying the remaining 2.4 GHz bandwidth with 60 dB of isolation, at an estimated weight and power consumption of approximately 400 kg and 100 W, respectively.

ACKNOWLEDGMENTS

William D. Glenn generated the baseband processor architecture, and Thomas A. Schonhoff, Robert R. Pugh, and Sastri L. Kota consulted on modulation, technology, and system control, respectively. Ersch Rotholz contributed to the satellite switching concepts developed earlier under this contract. He was inadvertently omitted from the acknowledgments in the Executive Summary of the First Year Report. Mr. Rotholz and Paul F. Christopher also served as consultants during the present effort. Ruth W. Wales provided editorial assistance, and Helen E. King, Joan H. Johnson, Judy A. Laskey, and Elizabeth A. Godwin prepared the manuscript.

The authors are pleased to thank these individuals for their valuable contributions.

TABLE OF CONTENTS

<u>Section</u>		<u>Page</u>
LIST OF ILLUSTRATIONS		viii
LIST OF TABLES		xi
1	EXECUTIVE SUMMARY	1
2	INTRODUCTION	5
3	BASEBAND PROCESSOR	9
3.1	BASELINE SYSTEM ASSUMPTIONS	9
3.1.1	Uplink and Downlink System Interfaces	12
3.1.2	Rationale for System Parameter Selection	13
3.2	FEATURES OF ON-BOARD PROCESSING	15
3.2.1	Signal Regeneration	15
3.2.2	Baseband Switching	16
3.3	DEMODULATION/REMODULATION	17
3.3.1	All-Digital Implementations	17
3.3.2	Hybrid Realizations	19
3.3.3	Microwave Demodulation	19
3.3.4	Implementation Considerations	22
3.4	NETWORK PROTOCOLS	23
3.4.1	Link Establishment Protocol	23
3.4.2	Packet Routing Protocols	31
3.4.3	Protocol Time-Out Parameters	46
3.5	PACKET FORMATS	49
3.5.1	Control Packet Format	50
3.5.2	Message Packet Format	58
3.5.3	ACK/NACK Message Packet Formats	60
3.5.4	Short ACK/NACK Packet Formats	62
3.6	SIZING THE BASEBAND PROCESSING	64
3.7	THE PROCESSOR ARCHITECTURE	66

TABLE OF CONTENTS (Continued)

<u>Section</u>	<u>Page</u>
3.8 PROCESSOR OPERATION	69
3.8.1 Point-to-Point Transmission	69
3.8.2 Multipoint Transmissions	72
3.8.3 Implementation of the Baseband Processor	73
3.9 TECHNOLOGY IMPLICATIONS	89
3.9.1 Baseband Processor Data Rates	89
3.9.2 Major Areas Requiring State-of-the-Art LSI Technology	90
3.9.3 Overview of LSI Technology	92
3.9.4 Overview of Commonly Available State-of-the-Art	131
3.9.5 Recommendations	140
3.10 CONCLUSIONS/RECOMMENDATIONS	141
3.10.1 Baseband Processor Interfaces	141
3.10.2 Bit Processor/Packet Switch	142
4 MICROWAVE SWITCH	147
4.1 MICROWAVE SWITCH ASSEMBLY CARD ORGANIZATION	149
4.2 CROSSPOINT CONFIGURATION	153
4.2.1 Three-Stage Amplifier	155
4.2.2 Dual Row/Column Board	158
4.3 MICROWAVE SWITCH CONTROL LOGIC CONFIGURATION	158
4.3.1 Memory Levels	161
4.3.2 Basic Operation	161
4.3.3 Level 2 Memory Update	164
4.3.4 Level 1 Memory Load	166
4.4 INTERNAL ELECTROMAGNETIC COMPATIBILITY	167
4.4.1 Harmonics of the Switching Pulse on the Signal Lines	167
4.4.2 Cross-Coupling Switching Line-to-Switching Line	171
4.4.3 Signal to Switching Line Cross-Coupling	173

TABLE OF CONTENTS (Concluded)

<u>Section</u>	<u>Page</u>
4.5 MECHANICAL DESIGN	174
4.5.1 Hardware Implementation	174
4.5.2 Microwave Drawer	182
4.5.3 Weight Estimate	182
4.6 CONCLUSIONS	182
REFERENCES	187
APPENDIX A - TRAFFIC MATRICES	189
APPENDIX B - EFFICIENT ACCESS SCHEME FOR NON-REAL-TIME DATA	195
APPENDIX C - ALL-DIGITAL DEMODULATION	202
APPENDIX D - SAWD AND CCD SIGNAL PROCESSING OVERVIEW	212
APPENDIX E - BANDWIDTH AND POWER EFFICIENT MODULATIONS	220
BIBLIOGRAPHY	235
GLOSSARY	239
DISTRIBUTION LIST	245

LIST OF ILLUSTRATIONS

<u>Figure</u>		<u>Page</u>
3-1	Fundamental Baseband Processing Configurations	18
3-2	Approximate Operating Ranges of CCDs and SAWDs	21
3-3	The Link Establishment Protocol (Point-to-Point)	26
3-4	The Link Establishment Protocol (Multipoint)	29
3-5	The Link Establishment Protocol (Partial Multipoint)	30
3-6	The Packet Routing Protocol (PRP-1)	33
3-7	Redundant Packet Transmissions Using PRP-1	37
3-8	Packet Routing Protocol (PRP-2)	38
3-9	Packet Routing Protocol (PRP-3)	40
3-10	Packet Routing Protocol (PRP-4)	42
3-11	Packet Routing Protocol (PRP-5)	44
3-12	Control Packet Format	51
3-13	Message Packet Format	59
3-14	ACK/NACK Message Packet Format	61
3-15	ACK/NACK Packet Format	63
3-16	Simplified Block Diagram of the Baseband Processor	67
3-17	The Bulk Memory Controller	74
3-18a	The Bulk Memory Array (Input Side)	77
3-18b	The Bulk Memory Array (Output Side)	80
3-19a	The FIFO Controller	81
3-19b	FIFO Controller (Point-to-Point Address Decoder)	83
3-19c	FIFO Controller (Multipoint Address Code (MAC) Decoder)	84

LIST OF ILLUSTRATIONS (Concluded)

<u>Figure</u>	<u>Page</u>
3-19d FIFO Controller	86
3-20 A Single Switching Processor	88
3-21 DOD VHSIC Chip (0.5 μ m) Lithograph Goals (Contrasted) With Some Current Technologies	119
3-22 Yearly Trend of LSI Chip Density Increase	122
3-23 Comparison of Yearly Growth Trends in LSI Chip Density for Logic & Memory Devices	123
3-24 Yearly Improvement in Access Time for Fixed-Size (1K) MOS RAM	124
3-25 Yearly Improvement in Access Time for Several Sizes of Bipolar RAM's	125
3-26 Multiplier Requirements for 100 MS/S Data Rate	136
3-27 Three (Slower) Multipliers Pipelined to Maintain Sampled Data Rate	137
3-28 Status of A/D Conversion Development (1980)	139
4-1 Crossbar Switch	150
4-2 Snaking Concept for Switch Cards	151
4-3 Verticalization for Switch Cards	152
4-4 Row Board Cross Point	154
4-5 Column Board Cross Point	156
4-6 MESFET Amplifier Circuit Diagram	157
4-7 MESFET Amplifier Microwave Layout	159
4-8 Dual Card Layout Study	160
4-9 Second Level Memory Organization	162

LIST OF ILLUSTRATIONS (Continued)

<u>Figure</u>		<u>Page</u>
4-10	First Level Memory Organization	165
4-11	Signal Pulse of Control Line	168
4-12	Circuit of FET Gate	169
4-13	Model for Signal to Switch Coupling	175
4-14	System Clock	176
4-15	Level 2 Memory Configuration	177
4-16	Bi-directional Memory Bus Separations	178
4-17	Level 3 Memory	180
4-18	Decoders and Cross Point Cards in Drawer	183
A-1	CONUS Traffic Regions	191
B-1	Typical MSAP Satellite Channel Occupancy	196
B-2	Queueing Delay for MSAP Satellite Channel in Packet Slots of Duration 2.048 s	199
B-3	MSAP Occupancy for Terminals in Scanning Beam Coverage Areas	200
D-1	Generation of MSK	214
D-2	Typical 1 GHz Oscillators	215
D-3	Reflective Array Compressor (RAC) Example	216
E-1	Smoothing for L=0, 1, 2, 3; Baseband Phase and Frequency Pulses	225

LIST OF ILLUSTRATIONS (Concluded)

<u>Figure</u>	<u>Page</u>
E-2 Pair of 2D Smoothed Phase Trajectories Showing Underlying Phase Changes and Differences	229
E-3 Bound Parameter R_q vs E/N_0 for $M=4$ Jump Choices and Modulation Index $r/q = 1/4$ (Jump Increment $2\pi r/q = \pi/2$) Showing Effect of OD, 2D, and 3RC Smoothing	231
E-4 Bandwidth and Power Efficiency for 1D and 3RC Smoothing, $M=4$ Jump Choices, and 20 and 40 dB Bandwidths	233

LIST OF TABLES

<u>Table</u>	<u>Page</u>
3-1 Traffic Matrix Based on CONUS Customer Premises Terminal Population	11
3-2 Qualitative Comparison of SAWDs and CCDs	20
3-3 Connectivity for the Receiver/Transmitter Channels to the Switching Processors	68
3-4 Trade-off Between GaAs IC Technologies	103
3-5 Comparison of Ring Oscillator Performance for the Main GaAs Logic Families	103
3-6 Commercially Available Magnetic Bubble Memory Chips	116
3-7 DoD VHSIC Silicon Technology Goals	118
3-8 MOS: Current Capabilities and Goals	120
3-9 Technology Choices for VLSI	121
3-10 IC Device Performance Projections: 1985-1990	127
3-11 Typical Commercially Available 16-Bit Micro-processor Chips	132
3-12 Typical Commercially Available DMA Controller Chips	133
3-13 Typical Commercially Available LSI Data Encryption Chips	134
3-14 Typical Commercially Available Fast (Accumulating) Multipliers	135
3-15 Typical Commercially Available Fast S&H and ADCs	138
3-16 Power Estimates for Baseband Processor	143
4-1 Comparison of Logic Technologies	164
4-2 Margin of Signal over Control Pulse Harmonics	171

LIST OF TABLES (Concluded)

<u>Table</u>	<u>Page</u>
4-3 Sizing and Power Requirements	181
4-4 Weight Estimate	184
A-1 Traffic Matrix Based on CONUS Premises Terminal Population	190
C-1 Minimum Number b of Quantization Bits Required	204
C-2 Upper Bound on Total Power (W) Consumed by b x b GaAs Multipliers for W = 100 MHz and N = 16	206
C-3 Approximate Mini-rotation Angles and Gain Constants	208
C-4 Rules for Selecting (I ₁ , Q ₁) Given r ₋₁ , r ₀ , and (I, Q)	210

SECTION 1

EXECUTIVE SUMMARY

In the first year of this two-year study of on-board processing concepts for future satellite communications systems by the MITRE Corporation, we identified an exemplar 30/20 GHz system configuration. We suggested cost-effective directions for further consideration which indicate a variation of about 30% in total system cost as a function of overall system design. Key aspects of the system are the multibeam antenna and utilization of the 2.5 GHz bandwidth. We estimated that 100 MHz of spectrum would be utilized for the customer premises applications served by two scanning beams, and 2.4 GHz for the fixed users served by 40 fixed beams.

In order to provide numerical discussions for the second year study of the regenerative baseband processor, it was assumed that there were eight fixed and eight scanning beams on both the uplinks and downlinks, each using the same 100 MHz uplink and downlink bandwidths. The estimated power consumption is 1 kW. With 1 b/s per hertz and 16 movable spot antenna beams, the system would have a capacity of 1.6 Gb/s, which is considered sufficient to emphasize processor design concepts and device technology appropriate to an advanced concept study. For the fixed users employing the 2.4 GHz bandwidth, a separate 100 x 100 crosspoint radio frequency (RF) switch design was studied.

BASEBAND PROCESSOR

The second year study of an all-digital baseband processor produced a new processor architecture and provided more detailed discussions of modulation, link establishment, network and packet routing protocols, and packet design. These features are closely related to the design of the processor as well as to throughput efficiency and performance requirements of the technology.

Memory requirements were reduced by adopting a 1 ms frame duration of eight 125 microsecond time slots. This design also allows multiple frequency division multiple access (FDMA) carriers with time division multiplex (TDM) in a fixed beam, or a number of FDMA signals using a single channel per carrier during the dwell interval of a scanning beam. With the beam scan occurring once per frame, the burst rate is 8R for a data rate R, and also 8R for time

division multiple access (TDMA) users in a fixed beam. With R equal to the rate of a T1 carrier (1.544 Mb/s), $8R$ is 12.3 Mb/s and eight channels use 100 MHz (one bit per cycle or 1 b/s per hertz).

Digital demodulation requires analog to digital conversion followed by high speed multiplication. Various technologies can be utilized but none are developed adequately for the capacities envisioned in this system. Gallium arsenide (GaAs) very large scale integration (VLSI) technology appears feasible for all-digital demodulation with further development impetus. Hybrid analog techniques using surface acoustic wave and charge-coupled devices were investigated, and also are recommended as serious candidates for demodulators.

Demodulated data can be flexibly routed using packet techniques. A link establishment protocol utilizing demand assignment schemes is used to establish links and packet routing protocols (PRPs) are employed for message transmissions. The two levels of protocol are worked out in detail, including five different PRPs. The associated packet formats are also detailed.

RF CROSSBAR SWITCH

For the high capacity trunking applications, the 100×100 switch architecture (microstrip directional coupler with field effect transistor (FET)) switches identified in the first year report was carried a step further toward a physical realization, and may be viewed as a more detailed organization of the switch design problem.

The switch can be visualized as two parallel planes of 100 parallel transmission lines which cross with lines of each plane in orthogonal directions: each crossing represents a crosspoint which requires a directional line-to-line coupler and FET switch. This large planar assembly was subdivided into cards with card-to-card cabling to maintain interconnectivity. A packaging concept of ten drawers of 20 cards each was chosen.

The critical microwave design of the crosspoints was given attention. A wiggly-line coupler having 2.4 GHz bandwidth realized in microstrip has the required characteristics.

A three-state FET is used as a switch. Considerable attention was paid to the design and mounting of logic and its interconnection in order to minimize cabling weight, provide adequate electro-magnetic compatibility (EMC) isolation, low cross coupling of

control lines, and adequate (60 dB) crosspoint isolation when a switch is open.

Switches are controlled from a memory pattern with changes in state at timed intervals. Memory is organized in three levels: (1) storage of changes in switch state for frame period, updated from ground command; (2) main control memory, pattern or menu; and (3) holding registers for each switch. Minimum weight and size occur when memory is distributed to locations close to the crosspoints.

Actual hardware was chosen to implement logic functions in order to illustrate feasibility. The hardware design represents a benchmark of feasibility and is suitable for VLSI consideration. Similarly, the latest microwave research in monolithic circuits allows weight reduction.

Estimated weight for the 100 x 100 switch is 387 kg. The weight was estimated by weighing some integrated circuits (ICs) and other parts, but over 80% was calculated on the basis of simplistic mechanical design. Consequently, a 30% tolerance must be assumed. We believe that the present design exercise has provided further credibility to the recommended approach, but that weight reduction is possible.

Estimated power consumption is 107 W. The 5000 double flip-flops used to control the 10,000 switch crosspoints consume 100 W of this power.

SECTION 2

INTRODUCTION

The anticipated growth in demand for telecommunications services will accelerate with the introduction or elaboration of new services (such as video conferencing and high speed computer networks). These new services together with the continuing growth in private line, video broadcast networks, and other familiar services lead to the conclusions that satellite communications will have an expanding role and that United States domestic satellite capacity achievable in the 4/6 and 12/14 GHz bands will not be sufficient in the 1990 time period. Since satellite communications have proven cost-effective for domestic telecommunications, expansion of the satellite capability into the 1990's appears essential. Accordingly, National Aeronautics and Space Administration (NASA) Lewis Research Center is conducting a program leading toward development and demonstration of technology for a 1990's satellite system which can expand the domestic communications capacity available from the geostationary arc.

The NASA Lewis program has focused on the development of a new frequency band (30/20 GHz) with frequency reuse as a central feature. This approach will provide technology for greatly extending the utility of U.S. domestic satellite communications in the 1990's and beyond. System concepts devised to provide frequency reuse and high capacity incorporate multiple beam uplink and downlink antennas interconnected by either demodulate/remodulate processing (in which the signals are demodulated to data bits and reformed for downlink transmission) or RF switching (in which the signals are not demodulated). Neither concept is strictly new, but for the applications envisioned, considerable extension of the state-of-the-art is implied. The MITRE Corporation has completed a two-year study of on-board processing for NASA Lewis Research Center as an integral part of the 30/20 GHz program. This volume is the final technical report of the second year study.

The first year of the MITRE study considered the system context for on-board processing and provided an exemplar system description which incorporated both demodulate/remodulate digital baseband processing and RF switching. The relationship of signal waveforms to efficiency of spectrum utilization and demodulation demands required a study of modulation. In addition, network control protocols were analyzed in relation to baseband processing and system throughput. Particularly emphasized were basic architectures for large scale RF switches and baseband processors, and an analysis

of system performance improvements realized from processing. The first year study resulted in three reports: a detailed technical report[1], an annotated bibliography [2] listing the results of a thorough literature survey which located 272 papers and reports, and an executive summary. [3]

The first year results concluded that significant performance gains on the order of 4 dB in energy to noise ratio (E/N_0) resulted from demodulate/remodulate processing, and pointed towards a microprocessor all-digital approach for further analysis.

Investigation of RF switch architectures led to identification of a recommended approach: a microwave crossbar realized in microstrip with feedthrough directional couplers and FETs as switching elements.

These concepts (digital baseband processor and RF crossbar switch) were further analyzed in the second year and results are reported in this volume. Included is a more detailed and up-to-date treatment of digital technology for the baseband processor, and more detailed discussions of modulation, all-digital demodulation, network protocols relating to efficient access, and packet routing in the baseband processor. An engineering analysis of the RF switch showed how the hardware could be packaged and what power and weight would result by using existing technology.

The baseband processor design is not built directly on a NASA traffic model because the regenerative portion of the satellite is severely limited by foreseeable technology. For example, only about 100 MHz of the available 2.5 GHz bandwidth is feasible for baseband processing. However, most results of this study will apply to any traffic model. Traffic model discussions are found in Appendix A. These considerations provide a workable foundation for baseband processor designs. The system presented in this report is for illustrative purposes and does not necessarily suggest an optimum solution. Technology improvements and/or subsequent optimum designs will provide greater capability for handling more demanding traffic model requirements at baseband.

The satellite is organized so that users may send their low data rate traffic to the baseband processor which will demodulate, process, route and remodulate the users' messages. Users requiring higher data rates will use the microwave switch on-board the satellite rather than the on-board baseband processor. The microwave switch accomplishes message routing by switching the undemodulated RF carriers from an uplink beam to a selected downlink beam. Section 3 of this report deals with the baseband processor while section 4 covers the microwave switch. A more detailed description of each section follows.

Section 3 begins with a brief introduction and a definition of what the baseband processor entails. The baseline system assumptions concerning the uplink and the downlink system interfaces and the rationale for system parameter selection are covered in 3.1. Signal regeneration and baseband switching, features of on-board processing, are discussed in 3.2. Section 3.3 deals with several topics concerning demodulation and remodulation of signals. The link establishment protocol and the packet routing protocols required to maintain user synchronization with the baseband processor are presented in 3.4. Section 3.5 provides a description of the packet formats needed to support the various network protocols. The baseband processor is sized in section 3.6 and its architecture is investigated in 3.7. The operation of the processor during point-to-point transmissions and multipoint transmissions is explained in 3.8. Section 3.8 also includes a look into possible implementations of the baseband processor. Technological implications of the processor are explored in 3.9. This section provides a review of the processor data rates, a look at major areas requiring state-of-the-art LSI technology, an overview of LSI technology, an overview of commonly available state-of-the-art large scale integration (LSI) products, and recommendations. Conclusions and recommendations concerning the baseband interfaces and the bit processor/packet switch are presented in 3.10, finishing the documentation of the baseband processor.

Section 4 begins with a brief introduction and description of the microwave switch portion of the satellite. The microwave switch assembly and organization is explained in 4.1. A look at the crosspoint configuration is provided in section 4.2. The various memory levels and the basic operation of the microwave switch control logic are explained in 4.3. The problem of internal electromagnetic compatibility is investigated in 4.4. This investigation includes a look into the harmonics of the switching pulse on the signal lines, the cross-coupling switching line-to-switching line and the signal to switching line cross-coupling. A presentation of the mechanical design of the microwave switch, including its hardware implementation and a weight estimate, is found in section 4.5. Section 4.6 contains conclusions about the microwave switch based on this research.

This report contains several appendixes that provide a more detailed investigation of many satellite communications related topics. Appendix A contains a traffic matrix based on a postulated customer premises terminal population. This appendix explains the matrix derivation and provides suggestions of how the matrix might be used in conjunction with the satellite communications system under study.

An efficient access scheme for non-real-time data is presented in Appendix B. Low duty factor terminals with non-real-time data can utilize satellite channels very efficiently without any centralized control. This appendix addresses the feasibility of this concept for small user as an alternative to the use of the network protocols.

Two approaches to all-digital demodulation schemes are considered in Appendix C. The first approach investigated is the straightforward Analog-to-Digital (A/D) conversion of the uplink signals followed by the multiplication and accumulation of b-bit quantized samples. The second approach utilizes the CORDIC rotation.

An introductory overview of surface acoustic wave devices (SAWDs) and charged coupled devices (CCDs) is presented in Appendix D. References are provided to the reader interested in further details.

Appendix E updates and expands on material that appeared in MITRE's first year reports to NASA [1, 2] concerning bandwidth and power efficient modulations.

SECTION 3

BASEBAND PROCESSOR

In a narrow sense, a baseband processor is the subsystem of the satellite communication payload that manipulates bits (b) or sequences of bits (packets) of information gleaned from the digital signals received on the uplinks. The inputs and outputs of this subsystem are binary data only. The input bits are derived from incoming intermediate frequency (IF) bandwidth signals by the process of demodulation which converts individual signals to their relatively smaller information bandwidths (basebands) through detecting the bits of data imparted to the transmitted signals. The output bits are the basis for the remodulation process whereby uplink signals are regenerated for the satellite downlinks.

According to this definition the baseband processor hardware is purely digital in nature. However, since many functions of packet buffering, reformatting, switching, etc., are performed, considerable software or firmware is also associated or included with the baseband processor. In this section, the baseband processor hardware aspects are covered by discussions of a multi-microprocessor architecture and candidate implementations using state-of-the-art or predicted electronic technologies. The software aspects are treated primarily in terms of packet protocols.

In a wider sense, the baseband processor includes the demodulators and remodulators, because they constitute the interfaces through which the on-board digital processor communicates with the outside world. The uplink interface is emphasized in this section since the demodulators potentially have the greatest impact on communications payload weight and power. Bandwidth and power efficient digital modulations are also highlighted.

A baseline system design resulting in a baseband throughput capacity of approximately 1.6 Gb/s or 1.024 M, 1544 b packets per second(s) is postulated below. This provides a specific context for more detailed subsystem design and analysis.

3.1 BASELINE SYSTEM ASSUMPTIONS

The essential ingredients of the advanced domestic satellite communication system are multiple beam antennas for frequency reuse, bandwidth efficient modulation and coding for spectral and power

conservation, and on-board processing and switching for servicing many different traffic and user requirements.

Satellite weight and power constraints which might be associated with baseband processing are expected to be in the order of 400 kg and 2 kW. This would include the functions of uplink demodulation, baseband processing of data packets, remodulation, and downlink transmission power; antenna weight is not included. With this weight and power budget, one can expect to support in the order of 100 MHz of spectral bandwidth devoted to signals to be processed at baseband. [1] The remainder of the 2.5 GHz bandwidth allocation at 30/20 GHz would be used for signals which are to be RF switched and not regenerated on-board the satellite. The frequency bands should be reused many times by utilizing multiple beam antennas.

A fundamental assumption for the perceived relative traffic flow among terminals and through this domestic communication satellite determines the average connectivities among users. The assumption is that communications traffic from area A to area B is directly proportional to the product of the numbers of customer premises terminals in region A and region B. This is contrary to the intuitive assumption pertaining to normal telephone traffic where completely localized conversations constitute a relatively fixed fraction of the total traffic of a region regardless of the population of that region. The traffic model adopted here is more expected for communications among private corporations, government agencies, universities, and other institutions employing rooftop terminals.

A further simplifying assumption is that the distribution of customer premises terminals is directly proportional to human population. Given a list of the more densely populated areas in the country and a corresponding distribution of customer premises terminals, a reasonable traffic matrix can be generated and used to estimate the baseband switching connections or slot allocations on-board the spacecraft. An example of such a traffic matrix produced by MITRE using a terminal distribution postulated by TRW [4] is shown in table 3-1. (See also Appendix A.) If the extreme entries in the traffic matrix are too different to reconcile with slot assignments alone, then perhaps the discrepancies can be partially recouped by power control which would permit different burst rates as well.

For users requiring high link availability, e.g., 0.999 and above, site diversity is required to overcome rain attenuation. Power control on the uplinks or downlinks might also be used to advantage.

Table 3-1
Traffic Matrix Based on CONUS Customer Premises Terminal Population

		8 areas/region (see Appendix A)																TOTAL	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16		
NY	LA	CHIC.	DET.	PHILA.	SAN FRANC.	WASH. D.C.	BOST.	REGION 1	REGION 2	REGION 3	REGION 4	REGION 5	REGION 6	REGION 7	REGION 8	REG.	REG.	TOTAL	
TMIS	700	505	505	321	306	228	220	.210	.1544	.1326	.923	.862	.849	.560	.310	.262	.9631		
TMIS AREA	700	505	505	321	306	228	220	.210	.193	.166	.115	.108	.106	.70	.39	.33			
1. NY		.0727	.0524	.0524	.0333	.0318	.0237	.0228	.0218	.0200	.0172	.0170	.0112	.0110	.0073	.0040	.0034	1	
2. LA		.0378	.0378	.0240	.0229	.0171	.0165	.0157	.0145	.0124	.0086	.0081	.0079	.0079	.0052	.0029	.0025	.7214	
3. CHICAGO		.0378	.0260	.0229	.0171	.0165	.0157	.0145	.0124	.0086	.0081	.0079	.0079	.0052	.0029	.0025	.7214		
4. DETROIT		.0153	.0146	.0146	.0109	.0105	.0100	.0092	.0079	.0055	.0051	.0051	.0033	.0018	.0016	.0016	.0016		
5. PHILA.		.0139	.0103	.0100	.0095	.0088	.0075	.0075	.0052	.0049	.0048	.0048	.0032	.0018	.0015	.0015	.0015		
6. SAN FRANC.					.0077	.0074	.0071	.0065	.0056	.0039	.0036	.0036	.0024	.0013	.0011	.0011	.0011		
7. WASH. D.C.						.0072	.0069	.0063	.0054	.0038	.0035	.0035	.0023	.0013	.0011	.0011	.0011		
8. BOSTON							.0065	.0060	.0052	.0036	.0034	.0033	.0022	.0012	.0010	.0010	.0000		
9. REGION 1									.0055	.0048	.0033	.0031	.0030	.0020	.0011	.0009	.0009	.2757	
10. REGION 2										.0041	.0028	.0026	.0026	.0017	.0010	.0008	.0008	.2368	
11. REGION 5											.0020	.0018	.0018	.0012	.0007	.0006	.0006	.1649	
12. REGION 6												.0017	.0017	.0011	.0006	.0005	.0005	.1539	
13. REGION 8													.0017	.0011	.0006	.0005	.0005	.1516	
14. REGION 3														.0007	.0003	.0003	.0003		
15. REGION 7															.0002	.0002	.0002		
16. REGION 4																.0002	.0468		

The existence of a large potential commercial market which is not being provided certain communications services is assumed. When these new services become a reality via future 30/20 GHz band satellites, the users may be content with more modest link availabilities, e.g., 0.99 and below, because they are not accustomed to any particular link quality and access expectations.

The economics of providing such service will not be focused on the present common carrier tariff structure; instead, the profit motive would focus more on the huge numbers of small terminals that could be sold in the satellite communications service catering to lower duty factor users. Availability of affordable terminals for a new service would create a technology pull, as opposed to demand push. Stimulation of technology-pull is appropriate for government action through a NASA development. Further pursuit of such philosophical issues is beyond the scope of this report.

3.1.1 Uplink and Downlink System Interfaces

In the baseline design, it is assumed that there are 16 independent uplink beams and 16 independent downlink beams associated with the baseband processor. Each uplink beam uses the same 100 MHz bandwidth in the uplink frequency allocation, and similarly on the downlink. Assuming a spectrally efficient modulation scheme which yields one b/s per Hz, the baseband throughput of the satellite will be approximately 1.6 Gb/s.

For the sake of specific discussion, the 16 uplink or downlink beams are arbitrarily divided into eight fixed beams and eight scanning beams. A common 1 ms frame interval, subdivided into eight 125 microsecond slots, is assumed for both types of beam. The beam and slot organizations permit a combination of FDMA and TDMA operation. In a fixed beam, a number of FDMA carriers, each employing TDMA within a frame, may be used. In a scanning beam, a number of FDMA signals operating in a single channel per carrier (SCPC) mode are present in the beam during a given dwell interval. In this case, each carrier is associated with a different terminal user. At the end of each 125 microsecond slot, a scanning beam is free to move to another coverage area. The period of every scanning beam is one frame or eight slots. Thus, nominally, a scanning beam would return to the same coverage area every eight slots, although the eight specific coverage areas visited could be reprogrammed from epoch to epoch. A fixed beam can be viewed as a special case of the scanning beam where the coverage area does not change.

In order to maintain a given data rate R , a terminal located in a scanning beam coverage area must burst its data at a channel rate

of $8R$. Similarly, in a fixed beam with a carrier utilizing TDMA, a given terminal must burst at $8R$. Consequently, if TDMA is not used on a fixed beam carrier, the burst rate is equal to R .

The nominal user data rate in the baseline system is 1.544 Mb/s. This is equivalent in data rate but not necessarily in signal structure to a T1 carrier. The baseline design will also incorporate T2 rate users at 6.312 Mb/s and lower rate users typically in the order of 64 kb/s.

A T1 rate user in a scanning beam must burst at approximately 12.3 Mb/s channel data rate during a scanning beam dwell. This, along with the one bit per cycle assumption implies that eight FDMA carriers at the burst rate may be accommodated in the 100 MHz bandwidth of the scanning beam. Note that in this sense, the one b/s per Hz assumption applies to the center frequency spacing between carriers. This implies that 64 T1 class users may be serviced concurrently within a scanning beam. With eight scanning beams and eight fixed beams, each interpreted as a special case of the scanning beam, a total capacity equivalent to 1,024 T1 users is implied.

3.1.2 Rationale for System Parameter Selection

It is important to understand the fundamental reasons for a given frame period and for the need for scanning beams. Although the particular values selected for the frame duration and the number of slots per frame are only nominal, some explanation of how they were selected follows.

It is assumed that the satellite is in either a geostationary or a near-geostationary orbit. This implies a round-trip propagation delay of approximately a quarter of a second. For real-time traffic, it is desirable that any satellite processing delay not contribute significantly to the overall delay.

One potential source of processing delay on-board the satellite derives from the possible accumulation of several bursts of uplink data from one source terminal in an FDMA mode to produce a single, higher rate burst of data to a single destination terminal in a TDM downlink mode. Assuming that each uplink burst is received in the same slot of the periodic satellite frame, and that the downlink burst will occupy the same slot duration, the key parameters are the frame duration and the number of frames required to accumulate the downlink burst data. In the most straightforward realization of the present approach, the ratio of the downlink burst rate to the uplink burst rate is equal to the number of slots in a frame. Thus, if an

uplink user has access to only one slot per frame and the satellite is accumulating data from that user for one particular downlink user, it takes a number of frames equal to the number of slots per frame to accumulate the entire downlink burst data. The parameters of 1 ms per frame and eight slots per frame imply a typical processing delay of 8 ms. Since this is much less than a one-way propagation delay, the integrity of real-time data will be preserved even though such on-board processing is performed.

Another form of on-board processing delay is that incurred by waiting for an appropriate downlink slot and/or beam for the destination of accumulated uplink data. This type of delay is typically upper bounded by the frame duration rather than some multiple of the frame period. Accordingly, this constraint is less stringent than the one implied by the FDMA/TDM conversion technique discussed above. For real-time data, the frame could be in the order of 10 ms long if it were determined just by the tolerable waiting time for downlink slots or beams. This would be possible if every packet received during an uplink slot always was retransmitted in a downlink subslot of one-eighth a slot duration within the next frame.

For non-real time data, much longer frames could be utilized. This could be advantageous for smaller, less expensive, low duty factor terminals which might employ ground-packet-radio type protocols with high channel efficiency. (See Appendix B.)

The number of slots in a frame was selected to be a power of two for convenience of binary representation in the baseband processor and in the order of ten to construct an illustrative system without too large a burst rate disparity between the uplinks and downlinks. The 1 ms frame is of the right order of magnitude to avoid excessive processing delay, and when subdivided into eight slots, a standard 125 microsecond slot duration compatible with T1 signal structures results. Uniform slot durations have been assumed in the baseline design for simplicity and to provide a convenient means of determining and controlling required burst rates and scanning beam dwell distributions. Although the resulting design would probably be more complex, it is certainly possible to envision non-uniform slot durations in conjunction with variable and programmable frame partitions.

The rationale for scanning beams follows from the basic system design objective of complete coverage of the continental United States (CONUS) and the highly non-uniform population density in this country. A reasonable engineering judgment for the baseline design is that a collection of fixed beams only would be an inefficient realization of spacecraft hardware and platform space. An

alternative recommendation is a combination of fixed beams and scanning beams where the scanning beams would be time-shared to cover the less densely populated and more remote regions of the country. The fixed beams would be designed for the metropolitan areas surrounding the major cities.

Each scanning beam would be preprogrammed to visit a given coverage area with a cumulative dwell time that is proportional to the population of that area. This schedule would normally be followed automatically on-board the spacecraft without any ground control. Perturbations from the average scanning beam dwell distribution would be specified by central ground command on a relatively infrequent basis to reflect gradual shifts in traffic demand. (See Appendix A.2.)

3.2 FEATURES OF ON-BOARD PROCESSING

The type of advanced domestic satellite envisioned for T1, T2 and lower data rate communications applications includes a regenerating repeater. This portion of the satellite demodulates uplink signals to baseband; processes the data bits by performing error control, baseband switching, and reformatting; and remodulates the data for the downlink. This contrasts with the RF switch portion of the satellite which does no demodulation or remodulation of T3 and T4 rate data but switches the signal channels among beams under a TDM scheme.

3.2.1 Signal Regeneration

The regeneration of data by the baseband processor has the advantage of improving performance or permitting smaller terminals and a lighter satellite to maintain a given margin. These reductions are due to lower power and smaller antennas and result in lower cost. Conventional transponder satellites simply hard-limit, frequency translate, and amplify the incoming signals. This repeats uplink noise and introduces on-board intermodulation products due to satellite nonlinearities. These degradations imply the need for higher power, larger antennas, and lower noise receivers to ferret out the useful signals.

On the uplink, the terminals can burst at much lower rates with FDMA than with TDMA. The burst rate and burst length difference factor equals the number of slots per frame. This difference can be significant in reducing terminal cost since it implies less effective isotropic radiated power (EIRP) for the same energy per bit and a more relaxed timing requirement.

Another issue is the degree to which terminals must be coordinated as members of a network with respect to transmission timing, frequency, and power. Many existing bandwidth efficient modulation schemes do not require stringent network coordination. In this case, the burden of demodulating simultaneous signals from independent terminals is placed on the satellite. Although this increases satellite complexity, it can simplify the terminals, and this philosophy may be economically attractive in a system of many more terminals than satellites.

Since signals are regenerated by demodulation and remodulation on-board the satellite, there is the opportunity for choosing a different downlink signal format than used on the uplink. In particular, TDM downlinks which may operate at many times the uplink burst rates, e.g., eight times in the baseline design, are recommended. In the FDMA mode of operation, several signals coming from separate earth terminals are simultaneously demodulated by the satellite. These data are switched, reformatted, and combined with data from other uplink beams into a set of TDM downlinks, one for each downlink beam. Each TDM downlink data stream is broadcast from the common satellite platform and received by all active terminals in the coverage region of that downlink beam. Destination terminals select portions of this TDM traffic addressed to them. This FDMA/TDM multiple access/multiplexing scheme leads to lower cost terminals and minimizes intermodulation problems of the satellite.

The fundamental reason for TDM rather than FDM downlink is the desire to avoid unnecessary intermodulation interference arising from generating too many downlink signals at distinct carrier frequencies with the simultaneous operation of nonlinear power amplifiers on-board the satellite. The higher downlink burst rate implied is more manageable in this case because all downlink signals emanate from the same platform, which may employ a common master timing reference. This makes the task of acquiring, tracking, and demodulating downlink bursts in the terminal much easier, compared to a situation where bursts from different terminals are generated from separate clocks (as on the uplink).

3.2.2 Baseband Switching

A much more important advantage of on-board processing in the form of satellite regeneration of the signal may be the great increase in flexibility for switching and interconnecting users on a bit or packet level basis. With the regenerating repeater, one is no longer restricted to a point-to-point network structure. One-to-many, many-to-one, and conferencing applications can be envisioned as well. By demodulating the lower rate data, opportunities for on-

board storage with buffers or bulk memory are introduced. This can be advantageous for collecting data to a destination not yet in a scanning beam coverage area, for example. It can also be useful for error control, link protocols, and reformatting data for the downlink, as noted.

The baseband switching topic is the heart of this section and is discussed at greater length in sections 3.4 through 3.8.

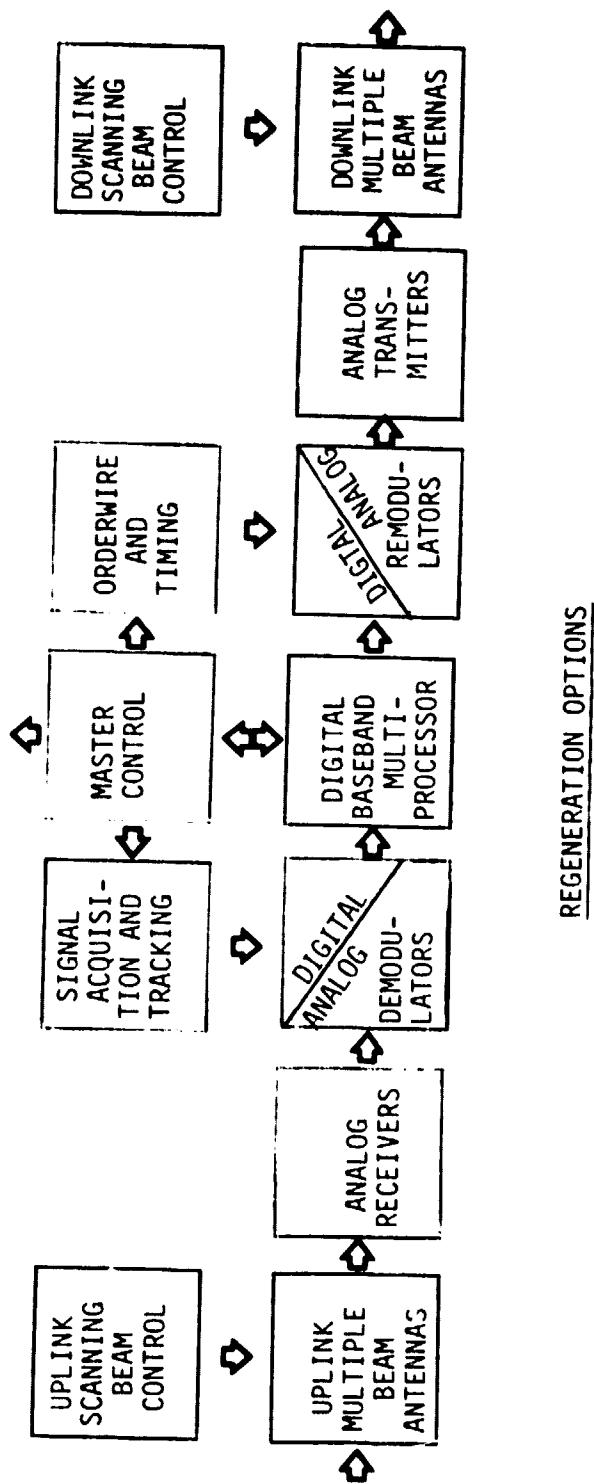
3.3 DEMODULATION/REMODULATION

Major contributions to spacecraft weight and power, resulting from the baseband processor and its communications related interfaces, are implied by on-board demodulation, especially if an all digital implementation is selected for the demodulators. Several alternatives for accomplishing the demodulation of uplinks, all digital, hybrid analog/digital, and all analog, are indicated in the general block diagram of figure 3-1. Each of these alternatives is discussed in some detail in an attempt to ascertain the most suitable approach.

3.3.1 All-Digital Implementations

The all-digital approach necessarily involves an RF mixing of the incoming signals to an IF bandwidth, followed by an analog-to-digital (A/D) conversion process with some level of input amplitude scaling and quantization, and subsequent purely digital arithmetic and accumulation operations assuming a strategy of rounding or truncation. The required high speed multiplications can be either performed with b-bit read-only memories (ROMs), b-bit high speed multiplier logic elements ($b \leq 5$, typically), or with CORDIC rotators. [5] These approaches are explained in detail in Appendix C. High power consumption of the digital circuitry required to perform demodulation of multiple signals in a single 100 MHz IF bandwidth may suggest that each signal be converted separately to its own IF bandwidth by an individual RF mixer.

The preferred alternative is the group demodulation of several signals within a sub-band after they pass through a single IF mixer and A/D converter. The state-of-the-art of A/D converters and power consumption calculations may suggest that this group demodulation approach is not feasible for the 100 MHz bandwidth of the baseline design. If it is feasible with respect to A/D conversion, there should be some advantages, including lower weight and power, to group demodulation as a compromise between simultaneous demodulation of completely synchronized signals using an (FFT) algorithm and separate demodulation of individual signals.



REGENERATION OPTIONS

- ALL DIGITAL
 - MULTIPLIERS
 - CORDIC ROTATORS
- ANALOG/DIGITAL
 - SAND/CCD HYBRIDS
- ALL ANALOG
 - MICROWAVE CIRCUITRY

Figure 3-1. Fundamental Baseband Processing Configurations

Signals with the same relative received power level could be allocated to the same sub-band according to either on-board processing or ground control protocols. This would result in either lower crosstalk, or permit the closer packing of the FDMA signals in the sub-band for a given crosstalk level. A closer packing of carriers would imply a narrower sub-band and the possibility of a lower A/D conversion sampling rate and less power consumption.

3.3.2 Hybrid Realizations

The hybrid method of demodulation may involve an analog/digital demodulation technique utilizing surface acoustic wave devices (SAWDs) and/or charged coupled devices (CCDs). With either family of devices, the basic signal processing and output detection involves analog quantities. (See Appendix D.) The relative advantages and disadvantages of SAWDs and CCDs are listed in table 3-2. This qualitative characterization should not be taken too literally because there is a large region of overlap in the bandwidth/time-delay space where either SAWDs or CCDs can be applied. (See figure 3-2.) Also, not all of the devices in either family possess every property shown in the table.

It is recommended that SAWDs and CCDs be examined as serious candidates for implementing spacecraft demodulators. Their main advantage is in the potential for low volume and power consumption. Although much has already been done in attempting to apply these analog devices to digital demodulation, additional development is warranted and on-going. SAWDs and CCDs, may represent a higher risk for spacecraft implementation than the all-digital approach for lower rate signals. However, the higher data rates of this application may make SAWDs and CCDs quite attractive for spacecraft implementation. The space qualification of these signal processing devices should have high priority.

3.3.3 Microwave Demodulation

The final alternative is a purely analog demodulation utilizing standard microwave circuitry.^[6] This approach has the outstanding advantage of being very low power because the devices themselves are passive. For standard digital modulation schemes like quadriphase shift keying (QPSK) and offset or staggered QPSK (SQPSK), these devices are conceptually very simple and presumably can be packaged into small volumes. On the other hand, this microwave approach may not lead to the same high density level of VLSI circuitry achievable

Table 3-2
Qualitative Comparison of SAWDs and CCDs

<u>Surface Acoustic Wave Devices</u>	<u>Charged Coupled Devices</u>
wideband (good for spread spectrum and combatting multiple access interference)	narrowband (can handle low data rates easily)
operate at IF or RF (mixers not required at VHF/UHF)	operate at baseband (mixers usually required)
passive but high insertion loss (non-volatile but amplifiers required)	active but no insertion loss (volatile and power required)
radiation insensitive (good for operation in near-earth space)	vulnerable to radiation (requires shielding in space)
may be temperature unstable (oven may be required)	temperature stable (good for space operation)
frequency may be imprecise (tuning may be required)	frequency stable (clocked operation)

Power, Size, Cost Can Be Comparable
 CCDs May Have Longer Mean Time Before Failure (MTBF)

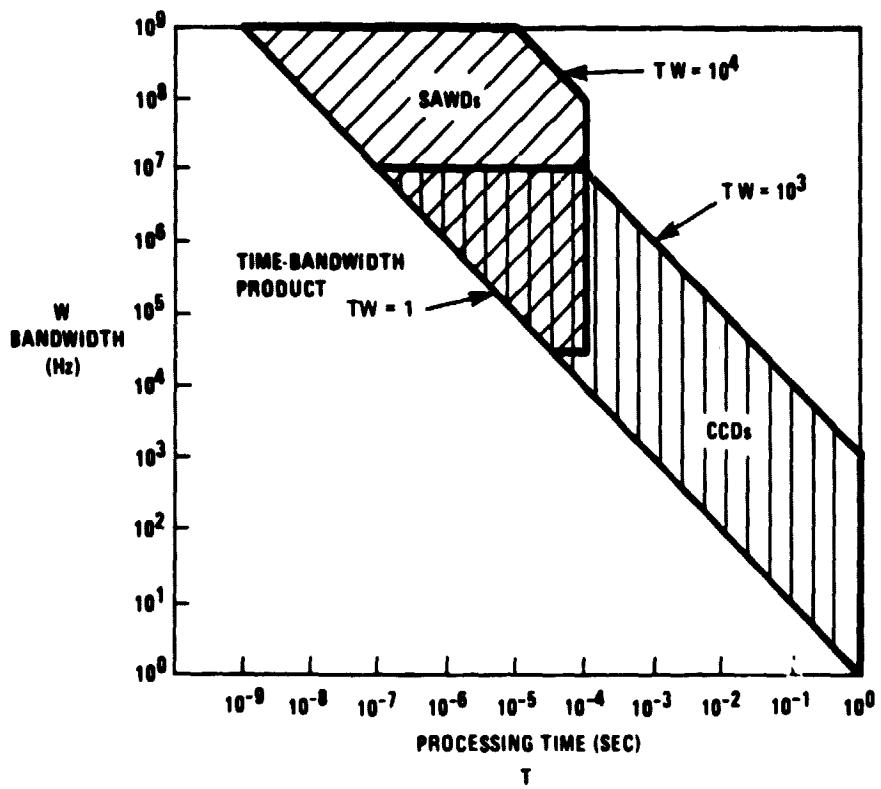


Figure 3-2. Approximate Operating Ranges of CCDs and SAWDs

with the all-digital approach or with SAWDs and CCDs. More importantly, the microwave circuits cannot be used with more advanced bandwidth efficient modulation schemes where the baseband modulation shape is not rectangular, as it is with QPSK. If there is a way of demodulating more sophisticated waveforms in a completely passive and analog fashion using microwave circuits, there does not seem to be any available literature on such techniques.

3.3.4 Implementation Considerations

Some of the more advanced modulation schemes include minimum shift keying (MSK), tamed frequency modulation (TFM), and various modulation schemes involving higher order alphabets, partial response signaling, and coding performance gains without bandwidth expansion. (See Appendix E.) These newer modulation schemes offer worthwhile improvements in both power and bandwidth efficiency which can be translated directly into a greater number of users being served and/or smaller weight and power on-board the spacecraft.

Since the most advanced modulations involve slowly varying phase transitions, and because this is so new, further development of signal acquisition, tracking and synchronization techniques is required. [7] Performance improvements with these modulations are obtained with greater complexity of signaling and demodulation. However, if the processing is done in a digital or hybrid fashion, as outlined above, increased performance may be worth the added complexity because of the continually improving higher density and lower power circuitry offered by VLSI technology.

There are further important tradeoffs to be made between standard modulation techniques with established implementations and more advanced modulation techniques which require greater complexity and challenge the ingenuity of circuit designers. Further investigations focused on these tradeoffs are strongly recommended. Only rough estimates of the power consumption for the all-digital approach are attempted in this report.

3.4 NETWORK PROTOCOLS

In order to design the baseband processor, several system requirements and goals must be defined. A major system requirement is the design of a network-wide protocol. This protocol must provide the users with a standard format in which to exchange messages, to maintain timing synchronization, to establish and terminate circuit connections, and to recover from message losses.

The network protocol under investigation in this report is organized as a two-level protocol. Each level is supported by a protocol which is custom tailored to meet the needs of the on-board processing satellite communications system. The protocol on the first level provides a demand-assignment multiple-access (DAMA) reservation scheme. The second level protocol controls the transmission of messages between the terminals. This two-level structure allows each protocol to operate independently.

The protocol independence achieved by this two-level scheme offers the network considerable flexibility. Updating or replacing an existing protocol on one level does not necessarily require an adjustment on the other level. As a matter of fact, it is possible to offer more than one level-two protocol to the users simultaneously. Users can then select the protocol best suited for their needs on a message by message basis. In addition, choosing one of the implementation options for one level does not reduce the number of options available for the other level. For example, the first level protocol, which is called the link establishment protocol (LEP), may be implemented either on-board the satellite or in a terrestrial control center. Selecting one implementation over the other has no impact on which second-level protocol may be used between the terminals. In the same manner, choosing one of the many available second level protocols, which are known as packet routing protocols (PRPs) will not limit the selection of level-one protocols. The PRPs may be organized as End-to-End protocols that are transparent to the satellite switching processing or as uplink protocols requiring on-board processor capabilities. Forward-error-correction (FEC) or automatic-repeat-request (ARQ) functions can also be supported by the PRPs. A look at the various PRPs available for satellite systems is presented later in this section. A review of the LEP is presented next.

3.4.1 Link Establishment Protocol

The primary function of the LEP is straightforward: the LEP must provide every user with the ability to request and to receive a channel assignment. In order to support this service, the protocol

must provide a standard procedure in which channel requests, assignments, and denials can be made in a format usable by all terminals. Since the network connects terminals that use different data rates, message types, and special services, the protocol must be flexible in design and general in operation.

The orderwire scheme considered in this work allows a user to request a channel for the duration of his message. In addition, the user specifies the message type, its length, priority and destination(s). He may also request special functions (e.g., broadcasting, conferencing, etc.) and/or special data rates. This flexibility allows every user to custom order a channel suited to his needs.

The LEP can be implemented either on-board the satellite or in a ground station control center. The following depicts the flow of control packets as a user attempts to reserve a channel when the orderwire processor is located on the ground.

The source first sends a LINK REQUEST packet to the control center via the satellite. When the request arrives at the center it is processed. First, the network's ability to supply a channel is checked. If the network cannot provide a free channel, the source is sent a LINK STATUS packet via the satellite informing him of the busy channels. If the network has an available link, the center sends a QUERY packet to the desired sink via the satellite. This packet informs the sink of the pending channel request and inquires into the sink's availability. The sink answers the query by sending a RESPONSE packet to the center via the satellite. Upon receiving the response, the center generates and sends a link status packet to the source. If the sink is free, the center sets up the circuit while the link status packet is on its way to the source. If the sink is busy, the link status packet informs the source, who may want to request the channel again at some later time. The scheme requires a double satellite hop in each direction which will result in link establishment times of no less than one second.

When the controller is located on-board the satellite, the operation of the LEP itself is identical to that described above. The difference is that this particular implementation requires only a single satellite hop in each direction. This permits a minimum link establishment time of half a second. In addition to better response time, the probability of an error occurring during transmission is reduced since each packet travels over one satellite link rather than two. These benefits of on-board processing require increased satellite weight, size, and power consumption since additional memory, processors and control logic will be required to implement the LEP.

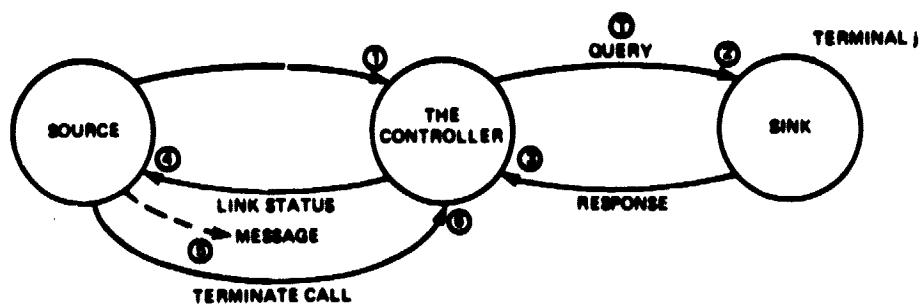
Although the function of this reservation scheme is to provide the users with a means of requesting contention-free message channels, the orderwire channel itself is not constrained to any particular accessing scheme. As a matter of fact, there are several accessing schemes which are suitable for use on the orderwire channel. The more popular schemes are Pure Aloha, Slotted Aloha and TDM.[8, 9] The selection of one accessing scheme over another for the orderwire depends greatly on the characteristics of the network.

Use of the Pure Aloha accessing scheme for the orderwire is recommended for networks that do not require frequent channel assignments. One such network would consist of a large number of terminals that transmit short, bursty messages on an infrequent basis. Another network suitable for the Aloha scheme is one with a large number of terminals that only need access to the orderwire processor after long periods of time because they are constantly busy transmitting long messages. Slotted Aloha could be used in similar networks to improve efficiency at the cost of additional complexity. In order for TDM to be efficient, the terminals must make use of their assigned slots as often as possible.

The reservation scheme proposed for this satellite communications system may not serve all users efficiently. Low duty terminals that transmit only a few packets at one time will incur a large amount of LEP overhead while reserving a message channel. These terminals should be provided an alternative accessing scheme. One possible alternative is to provide these users with a contention channel. Users could compete for this channel by directly transmitting their message packets. A Pure Aloha or Slotted Aloha scheme could be implemented on this channel.

The operation of the LEP begins when a user determines that he has a message to transmit. The flow of control packets due to a request for a POINT-TO-POINT circuit is shown in figure 3-3. A user (the SOURCE), generates a link request packet. Many different types of LEP CONTROL PACKETS exist. However, they are all organized within the same format to make processing easier. See section 3.5 for more information.

The link request is then sent to the orderwire processor (the CONTROLLER). This control packet notifies the controller that a user is requesting a circuit to a specific SINK in the network. The link request packet contains the address of both the source and the sink, as well as the length, type and priority of the message to be transmitted. In addition, the packet contains fields in which the source can specify special services such as non-standard data rates.



- ① THE SOURCE (TERMINAL i) TRANSMITS A LINK REQUEST TO THE CONTROLLER REQUESTING A LINK TO THE SINK (TERMINAL j).
- ② THE CONTROLLER SENDS A QUERY TO THE SINK AFTER RECEIVING THE LINK REQUEST.
- ③ THE SINK REPLIES TO THE QUERY BY SENDING THE CONTROLLER A RESPONSE.
- ④ IF THE SINK IS READY TO RECEIVE THE MESSAGE FROM THE SOURCE, THE CONTROLLER ESTABLISHES A CIRCUIT BETWEEN THE SOURCE AND THE SINK. THE CONTROLLER THEN SENDS THE SOURCE LINK STATUS INDICATING THE CIRCUIT HAS BEEN ESTABLISHED. IF THE SINK IS NOT READY, THE CONTROLLER SIMPLY SENDS THE SOURCE LINK STATUS INDICATING THAT NO CIRCUIT CAN BE PROVIDED AT THIS TIME.
- ⑤ IF THE SOURCE RECEIVES LINK STATUS PACKET INDICATING THAT THE CIRCUIT TO THE SINK IS READY, THE SOURCE TRANSMITS THE MESSAGE.
- ⑥ ONCE THE TRANSMISSION OF THE MESSAGE IS COMPLETED, THE SOURCE SENDS THE CONTROLLER A TERMINATE CALL PACKET. THE CONTROLLER THEN DISENGAGES THE CIRCUIT.

① = TIME-OUT CLOCK IS STARTED FOR THE PACKET.

18-59-775

Figure 3-3. The Link Establishment Protocol (Point-to-Point)

After the source transmits the request, he stores a copy of it in memory and starts a time-out clock. Should the request go unanswered beyond a pre-set time for any reason, the time-out clock will expire and alert the source to the problem. Whenever a time-out occurs the source increments the time-out counter associated with the lost packet. If the counter's value has not exceeded a specified number of tolerable failures, the packet is retransmitted and the time-clock is restarted. If the value of this counter exceeds a system specified value, the source enters a troubleshooting routine to determine the cause of the multiple transmission failures.

After receiving a control packet the first task the controller undertakes is the detection of errors. If the packet contains errors, the controller destroys the corrupted packet and moves on to a new task. Since packets are protected by time-out clocks, their destruction will result in either a retransmission or a maintenance check. When a control packet survives the error check, the controller determines what type of control packet has arrived.

If the control packet is found to be a link request, the controller checks to see what type of transmission mode has been requested. The network provides three modes for both data and voice messages. The standard mode is Point-to-Point which provides either a one-way or a bidirectional channel between one source and one sink. The Broadcasting mode option provides the source with a one-way channel to as many as 16 sinks. The third option is Conferencing. This mode provides a duplex channel between a source and as many as 16 sinks. With the address of the source and the sink(s), and the requested mode now known, the controller determines whether the channel hardware is available. If the system cannot provide the necessary hardware for the requested link, the source is notified via a link status packet which is sent to him by the controller. If the system can supply the link, the next task to be carried out by the controller is the determination of sink availability. This is accomplished by sending the sink(s) a query packet. These packets inform the sink of pending link requests and also inquire into the sink's availability.

A query packet is sent after the controller stores a copy of the link request packet that initiated the query packet, and the query packet's time-out clock is started.

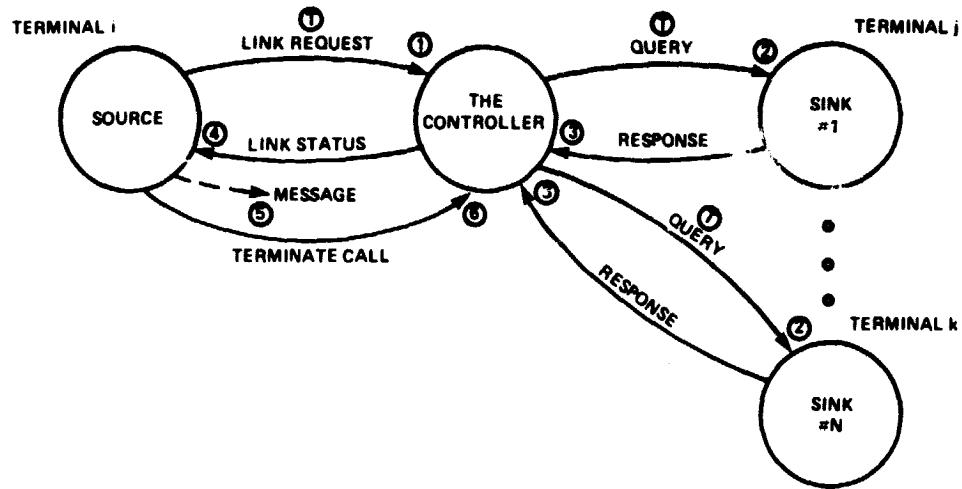
At the sink, the received control packet is checked for errors. If errors are detected, the control packet is destroyed. Error-free control packets receive further processing. After the query packet is processed, the sink generates a response packet based on the sink's current status. The response is then sent to the controller.

Upon arrival at the controller, the response packet is checked for errors. If the packet contains errors it is destroyed. This will cause the time-out clock associated with the original query packet to expire. The expired clock results in the retransmission of the query packet, which should result in the retransmission of the response packet.

If, however, the response packet was received error-free the controller clears the associated time-out clock. The controller then checks to see what type of connection the source and sink need. If the requested link is for point-to-point communications, the controller checks the response packet to determine whether the sink is available. If the responding sink is free, the controller generates and sends a link status packet to the source telling him the circuit is to be established. While the link status packet is on its way to the source the controller establishes the link. Whenever the requested sink for a point-to-point transmission is busy, the controller sends a link status packet informing the source of the situation. No link is established. The source must request the link again at a later time if he still requires the channel.

If the requested transmission mode is Multipoint (Broadcasting and Conferencing), the controller cannot send a link status packet until all the sinks queried reply. As each sink replies the corresponding link status packet is updated. Once the final response packet is received, the controller checks the link status packet that has been generated. If the status is that all the requested sinks are available, the controller establishes the links and sends the link status packet to the source. The source will begin transmission once the link status packet is received and correctly decoded. The flow of control packets when a Multipoint transmission mode is requested is displayed in figure 3-4.

If the controller discovers that not all the sinks are available, it must take a different course of action. This course of action is depicted in figure 3-5. The controller sends the link status packet to the source and it starts a time-out for this packet. The time-out is necessary since, unlike other link status packets, this one requires a reply from the source terminal in the form of a LINK ACCEPTANCE packet. The link acceptance packet is used to inform the controller whether or not the source wants the "partial link" to the available sinks. Thus, if the message can be sent to the unavailable sink(s) later, (e.g., one branch office sending electronic mail to the other offices), the source may decide to transmit to the available sink(s) first. If the sink(s) missing from the multipoint link is important, (e.g., teleconferencing where each board member called is required to discuss and vote on a business deal,) the source may not want to transmit to only those



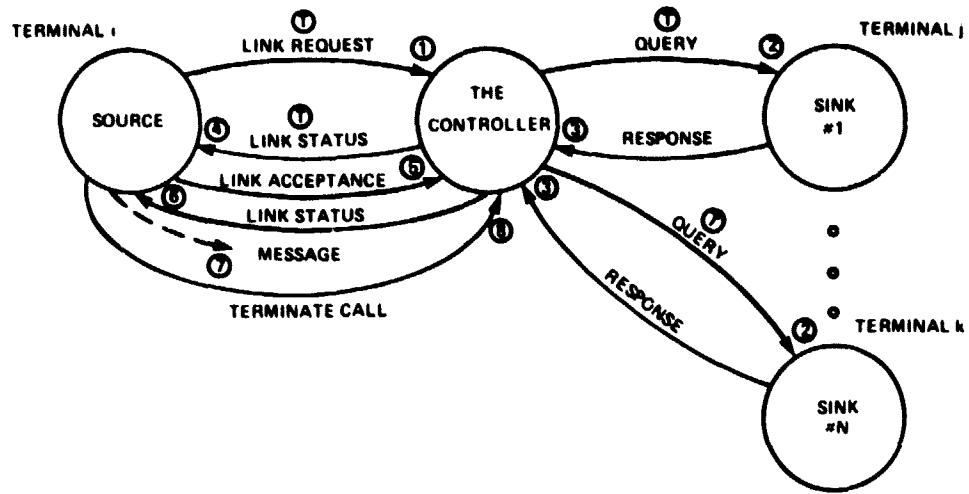
- ① THE SOURCE (TERMINAL i) TRANSMITS A LINK REQUEST TO THE CONTROLLER REQUESTING A MULTIPONT LINK TO SINKS 1 (TERMINAL j) THROUGH N (TERMINAL k).
- ② AFTER RECEIVING THE LINK REQUEST, THE CONTROLLER SENDS A QUERY TO EACH OF THE SPECIFIED SINKS.
- ③ EACH SINK REPLIES TO THE QUERY BY SENDING THE CONTROLLER A RESPONSE.
- ④ IF ALL THE SINKS ARE READY TO RECEIVE THE MESSAGE, THE CONTROLLER ESTABLISHES A CIRCUIT BETWEEN THE SOURCE AND THE SINKS. THE CONTROLLER THEN SENDS A LINK STATUS MESSAGE TO THE SOURCE INDICATING THAT THE CIRCUITS HAVE BEEN ESTABLISHED.
- ⑤ ONCE THE SOURCE RECEIVES THE LINK STATUS MESSAGE, TRANSMISSION OF THE MESSAGE BEGINS.
- ⑥ THE SOURCE SENDS THE CONTROLLER A TERMINATE CALL PACKET AT THE CONCLUSION OF THE MESSAGE TRANSMISSION.

⑦ TIME-OUT CLOCK IS STARTED FOR THE PACKET

N ≤ 15

IB-59.276

Figure 3-4. The Link Establishment Protocol (Multipoint)



- ① THE SOURCE (TERMINAL 1) TRANSMITS A LINK REQUEST TO THE CONTROLLER REQUESTING A MULTIPONT LINK TO SINKS 1 (TERMINAL 1) THROUGH N (TERMINAL K).
- ② AFTER RECEIVING THE LINK REQUEST, THE CONTROLLER SENDS A QUERY TO EACH OF THE SPECIFIED SINKS.
- ③ EACH SINK REPLIES TO THE QUERY BY SENDING THE CONTROLLER A RESPONSE.
- ④ IF NOT ALL OF THE SINKS ARE ABLE TO RECEIVE THE SOURCE'S MESSAGE, THE CONTROLLER SENDS THE SOURCE A LINK STATUS PACKET INDICATING WHICH SINKS ARE READY AND THOSE THAT ARE NOT. NO OTHER ACTION IS TAKEN AT THIS TIME.
- ⑤ AFTER RECEIVING THE LINK STATUS PACKET, THE SOURCE SENDS THE CONTROLLER A LINK ACCEPTANCE PACKET INDICATING WHETHER OR NOT THE SOURCE WANTS TO TRANSMIT TO ONLY THOSE SINKS CURRENTLY AVAILABLE.
- ⑥ THE CONTROLLER WILL ESTABLISH CIRCUITS FROM THE SOURCE TO THE AVAILABLE SINKS AND IT WILL ALSO SEND THE SOURCE A LINK STATUS PACKET INDICATING THIS HAS BEEN DONE ONLY IF THE SOURCE SENT A POSITIVE LINK ACCEPTANCE PACKET. IF THE SOURCE SENT THE CONTROLLER A NEGATIVE LINK ACCEPTANCE PACKET, NO CIRCUITS WOULD BE ESTABLISHED AND THE SECOND LINK STATUS PACKET WOULD NOT BE SENT.
- ⑦ ONCE THE SOURCE RECEIVES THE SECOND LINK STATUS PACKET, THE TRANSMISSION OF THE MESSAGE TAKES PLACE.
- ⑧ THE SOURCE SENDS THE CONTROLLER A TERMINATE CALL PACKET AT THE CONCLUSION OF THE MESSAGE TRANSMISSION.

- ⑨ TIME OUT CLOCK IS STARTED.

N ~ 15

IB-59-271

Figure 3-5. The Link Establishment Protocol (Partial Multipoint)

sinks currently available. Once a source makes a decision on whether or not to take the "partial link" he sends a link acceptance packet to the controller.

Upon receiving an error-free link acceptance packet (corrupt packets are destroyed) the controller either establishes the partial multi-point link or it does not, depending on the instructions from the source. If the links are established the source is sent a link status packet.

3.4.2 Packet Routing Protocols

When a user has established a link to his sink(s) via the LEP, he begins the transmission of his message. The exchange of messages between the terminals is controlled by the PRP. Since the network supports different users, data rates, message types and transmission modes, more than one PRP is needed to control the traffic. The proposed scheme requires the source to select the best suited protocol available from the network for each message he sends. The source must specify the protocol chosen when requesting the link so that the sink(s) can be notified. Thus, once the actual message transmission starts, both the source and the sink(s) are synchronized to operate under the same protocol.

Since the network may support many protocols, some users may not need to use every protocol. Thus, users could purchase only the software packages that they need (e.g., a user who will only transmit and receive voice messages would only need those PRPs which support voice communications). This scheme should be beneficial to the network as well as to the users. If the network is to be successful, a large number of users are needed to support it with their tariffs. In order to attract this large number of users, the network must provide services for all types of users. Potential customers who may want to join the network but lacked the initial capital needed to buy all the available software, could be enticed into the network with a STARTER PACKAGE: possibly a copy of the LEP software and one or two PRPs. Once in the network, they may expand their services and provide the network with additional tariffs. Larger customers could also be attracted to the network by its flexibility and by the large number of available software packages. The key to selling the network will be flexibility: flexibility in available services and in tariff rates.

This report investigates five packet routing protocols. The operation, features, disadvantages, advantages, and requirements of each protocol are presented. This survey is by no means exhaustive,

but it does provide an insight into the benefits and problems of a multi-protocol network (MPN).

- PRP-1

PRP-1 provides the network with an uplink ARQ scheme in addition to an end-to-end ARQ scheme. This protocol operates on a packet-by-packet basis. Figure 3-6a shows the normal flow of packets under the control of PRP-1. Figures 3-6b and 3-6c show the flow of packets under the control of PRP-1 during error conditions.

Operation of PRP-1 begins with the source first sending a message packet to the satellite. The source then stores a copy of this packet and starts a time-out clock. On-board the satellite, the received packet is processed by the switch. The packet header, which is protected by an error-correcting code, is corrected if necessary. Meanwhile, the packet body, which is encoded with an error-detection code, is checked for errors. If the packet contains errors, the source is sent a negative-acknowledgment (NACK) packet and the message packet is destroyed. The source must retransmit the message packet after receiving the NACK.

If the packet arrives at the satellite without errors, it is routed and sent to the proper sink. At the sink, the error-correction and detection routines are repeated on the packet. If any errors are detected in the packet body, the sink will disregard the packet and transmit a NACK to the source. The packet is accepted only if it is error-free. In this event, the source is informed of the successful packet transmission via an acknowledgment (ACK) packet sent by the sink.

As mentioned above, the source starts a time-out clock each time it sends a packet. The clock is reset every time an ACK or NACK is received. However, if the clock expires before the source receives either an ACK or NACK, the source assumes either that the packet was lost or the ACK/NACK reply was lost. The source can either re-transmit the message packet or (if the transmission failure rate has exceeded a system regulated level) enter a maintenance mode. The flow of packets in the event of a time-out is shown in figure 3-6d.

PRP-1 can be utilized for the transmission of short error-sensitive, non-real-time messages. This protocol may not be practical for long and/or real-time messages. The limitations of this protocol result from the long delays between sending a packet and receiving an ACK/NACK message. Due to the two round-trip propagation delays to and from the satellite, this delay is a minimum of 1/2 of a second. This does not include the delays due to

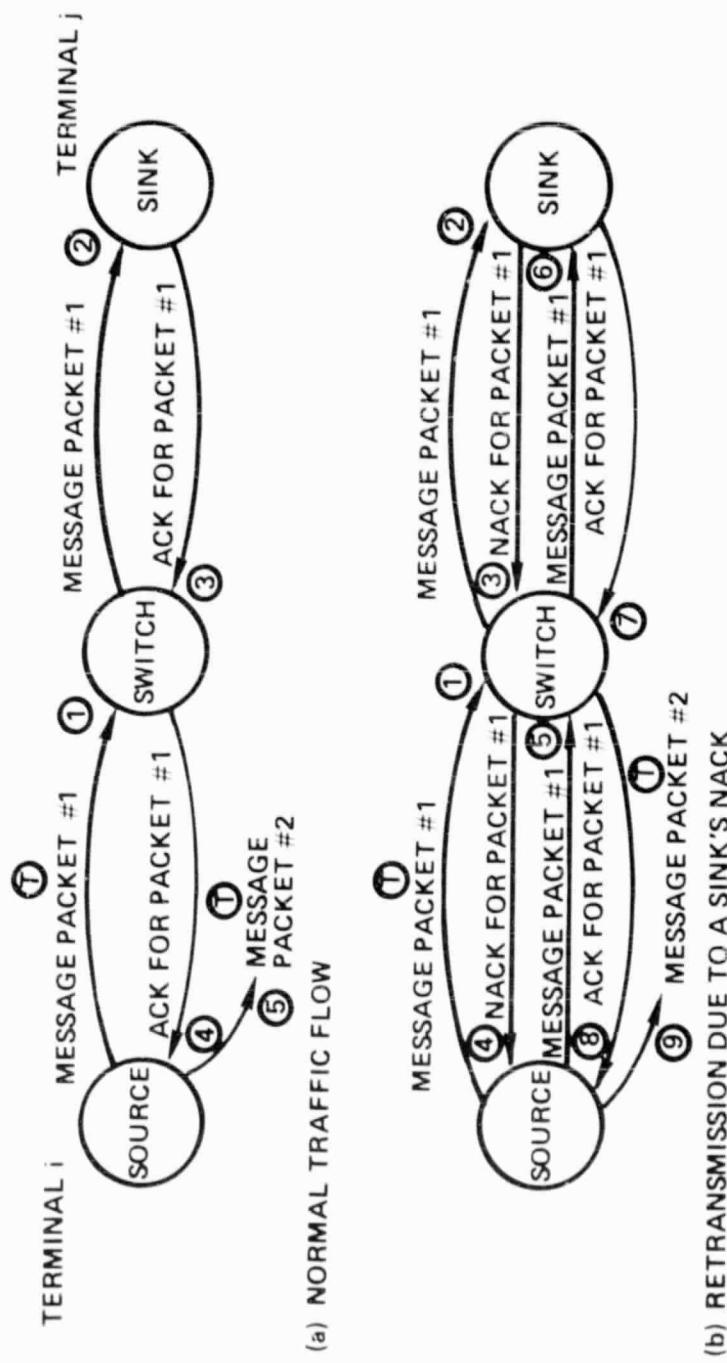
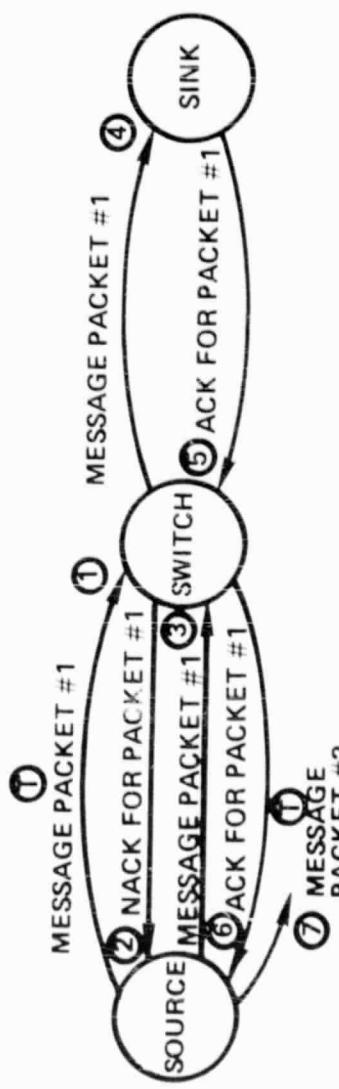
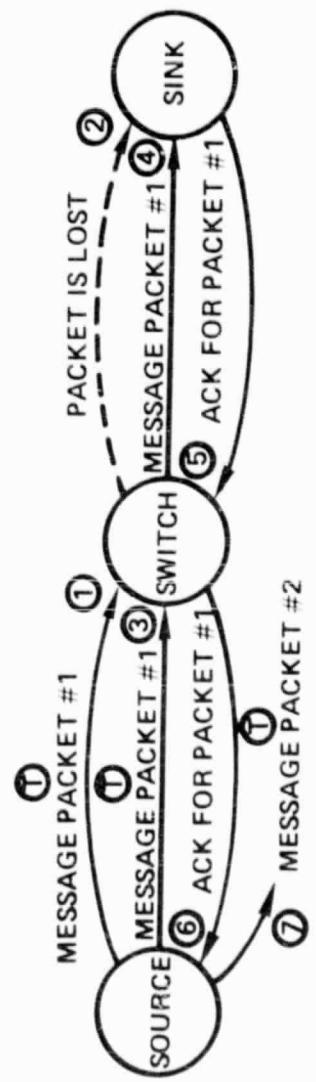


Figure 3-6. The Packet Routing Protocol (PRP-1)



(c) RETRANSMISSION DUE TO A NACK FROM THE SWITCH



(d) RETRANSMISSION DUE TO A TIME-OUT

Figure 3-6. The Packet Routing Protocol (PRP-1) (Continued)

processing times, routing, and waits in queues. Therefore, one can assume that the actual delay will generally exceed this minimum.

The following example is presented to illustrate why this particular protocol is not best suited for long messages. Assume a user has a message which consists of 10,000 packets; assume further that due to the current system loading, the delay time between transmitting a packet and receiving an ACK/NACK message is 3/4 of a second. Using PRP-1, the entire message transmission will require a total of 125 minutes to complete. Additional time will be required for every NACK received by the source since a retransmission is required. In addition to this long transmission time, the source is forced to store the large message for a long time. The source can only free one memory location every 3/4 of a second, provided an ACK was received for the packet residing in that location.

The major advantages of using PRP-1 are:

1. An uplink ARQ function is provided which reduces the consumption of downlink bandwidth by corrupted packets.
2. End-to-end ARQ is provided, ensuring the retransmission of packets received in error.
3. Time-out hardware and/or software is required only at the terminals. During transmission only the source is required to activate the time-out system.

The primary costs and disadvantages of using PRP-1 are:

- i. PRP-1 cannot be used efficiently for long messages.
2. Since packets may not arrive at the sink in real-time (~ 1 second), PRP-1 should be used only for non-real-time messages.
3. An error-correction capability is required at both the terminals and the satellite. Error-correction is a non-trivial problem for even modest size messages. However, since the protocol is based on the principle of ACK/NACK generation for each packet, it is essential that the packet headers be error-free. If the packet headers are not error-free, packets could be routed incorrectly, ACK/NACK messages could be sent for the wrong packet and/or not sent to the right source, and the retransmission of the wrong packets could also occur. The ACK/NACK packets themselves should be encoded with an error-correction code.

4. An error-detection capability at both the terminals and the satellite is required. In order for any ARQ scheme to operate, detection of errors is essential. Error correction does not pose as great a problem as error detection does. However, it still requires a careful investigation of the various schemes and system trade-offs.
5. As seen in figure 3-7, it is possible for the sink to receive duplicate copies of a valid packet. Therefore, the sink must have the ability to sort the incoming packets.
6. Smaller size ACK/NACK packets are needed since it is impractical to send full-size message packets for a single ACK/NACK. In order to use this smaller packet, additional hardware and/or software is needed at both the terminals and on-board the satellite.
7. All unacknowledged packets must be buffered by the source.

- PRP-2

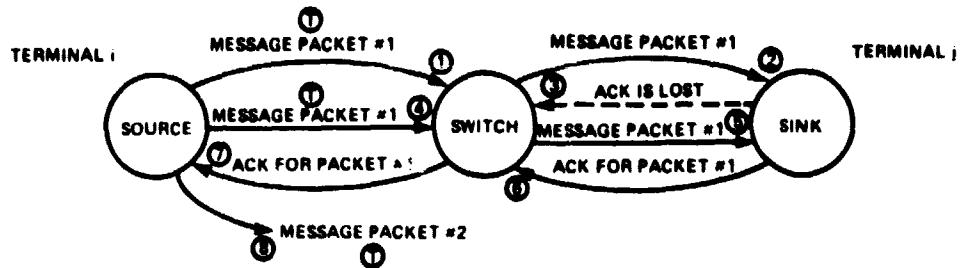
The operation of PRP-2 is shown in figure 3-8. PRP-2 is similar to PRP-1 with one exception; no uplink ARQ service is provided by PRP-2. The service that is provided, consists of an end-to-end ARQ scheme on a packet-by-packet basis. This protocol is virtually transparent to the switch since no protocol-related data processing is required of the switch. Like PRP-1, PRP-2 is best used for short, error-sensitive, non-real-time data messages.

The advantages of PRP-2 are:

1. End-to-end ARQ service is provided for each packet.
2. Time-out hardware and/or software is required only at the terminals. During a transmission only the source is required to use the time-out facilities.
3. Error-correction processing is not required on-board the satellite, although correction could be used to decrease routing mistakes due to packet header errors.

The costs and disadvantages of PRP-2 are:

1. PRP-2 cannot be used efficiently for long messages.

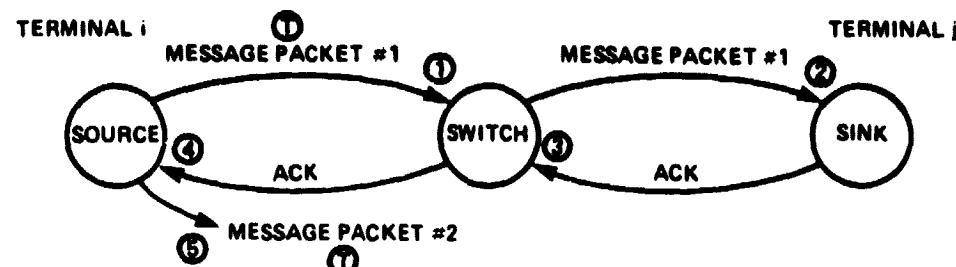


- ① THE SOURCE (TERMINAL 1) SENDS MESSAGE PACKET 1.
- ② THE SINK (TERMINAL 2) RECEIVES THE MESSAGE PACKET AFTER IT IS ROUTED THROUGH THE SWITCH.
- ③ AFTER PROCESSING THE RECEIVED PACKET, THE SINK SENDS AN ACK MESSAGE. HOWEVER THE ACK IS LOST IN TRANSMISSION AND IT NEVER REACHES THE SOURCE.
- ④ SINCE THE SOURCE NEVER RECEIVED AN ACK OR A NACK FOR MESSAGE PACKET 1, THE TIME OUT CLOCK EXPIRES. THE SOURCE THEN RETRANSMITS MESSAGE PACKET 1.
- ⑤ MESSAGE PACKET 1 ARRIVES AT THE SINK WITHOUT ERRORS. THIS IS THE SECOND VALID COPY OF MESSAGE PACKET 1 THE SINK HAS RECEIVED.
- ⑥ THE SINK SENDS ANOTHER ACK TO THE SOURCE.
- ⑦ THE SOURCE FINALLY RECEIVES AN ACK FOR MESSAGE PACKET 1.

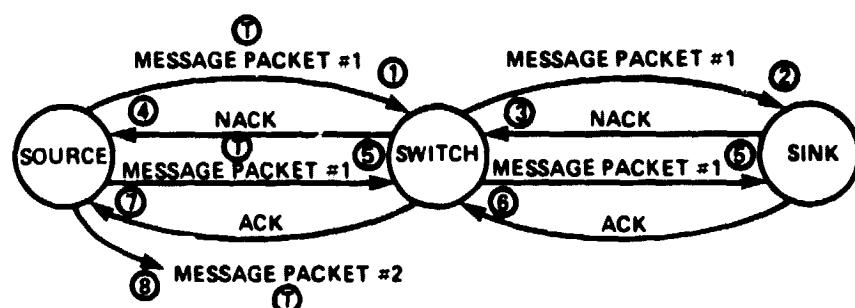
- ⑧ TIME OUT CLOCK IS STARTED FOR THE PACKET.

1859270

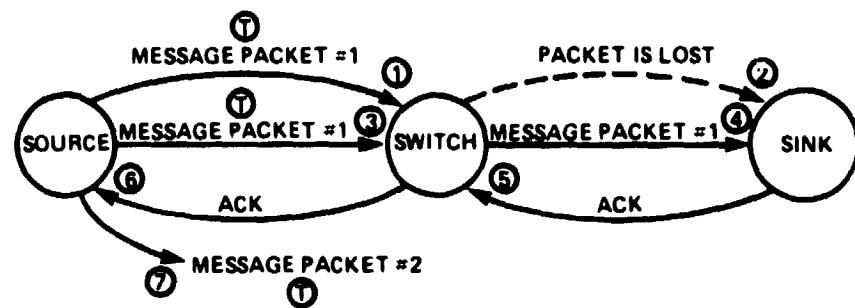
Figure 3-7. Redundant Packet Transmissions Using PRP-1



(a) NORMAL TRAFFIC FLOW



(b) RETRANSMISSION DUE TO A NACK



(c) RETRANSMISSION DUE TO A TIME-OUT

IA 59.267

Figure 3-8. Packet Routing Protocol (PRP-2)

2. Real-time messages are not readily supported.
3. Error correction is still required for the packet headers at the terminals.
4. Sinks may receive duplicate copies of valid packets.
5. Special size ACK/NACK packets are needed.
6. Downlink bandwidth is now being used to transmit packets that contain errors as well as the valid packets. Thus, some of the potential processing power on-board the satellite is untapped.
7. The source must buffer all packets that are unacknowledged.

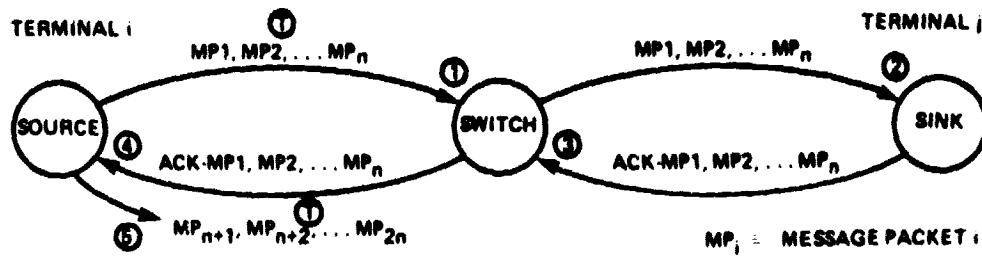
- PRP-3

The flow of packets under control of PRP-3 is shown in figure 3-9. This protocol allows users to handle their messages as strings of packets. When the source begins transmission to the sink, the source sends a fixed-length string of packets. The length of this string is selected during the link establishment procedure and it remains constant for the duration of the session. When the string of packets has been sent, the source starts a time-out clock and waits for an ACK/NACK message from the sink.

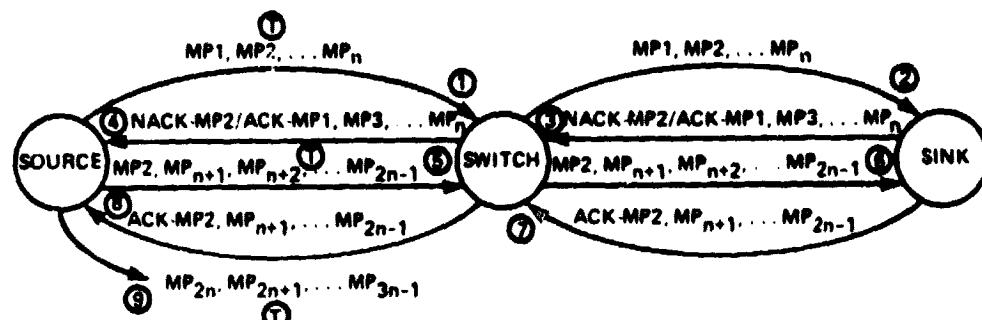
At the sink, the system receives and stores each incoming packet. Error correction is carried out on the header while error-detection is carried out on the packet body. The error status of each packet is included in a single ACK/NACK message for the whole string. When the final packet in the string arrives, the ACK/NACK message is sent to the source.

If the ACK/NACK message contains only ACKs, the source sends a new string of packets. Otherwise the source retransmits the packets identified by the NACKS. These packets are retransmitted in the next string, displacing some of the new packets that were slated for that string. If a message does not contain enough packets to fill an integer number of strings, the last string is shortened to the required length.

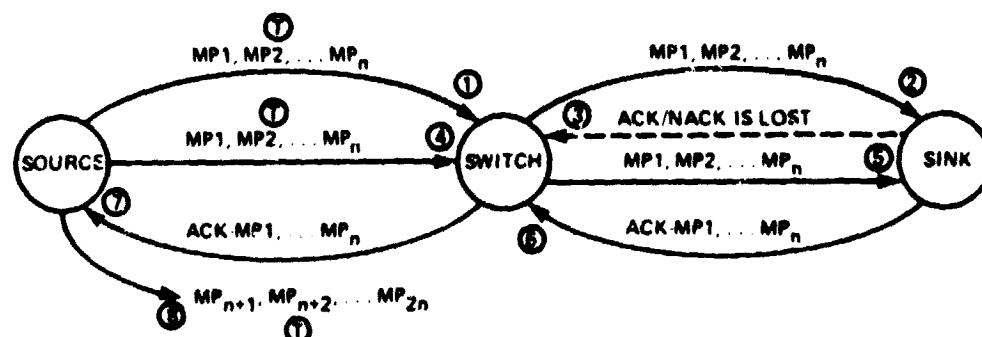
In the case of a time-out, the source is required to retransmit the entire string that has gone unacknowledged. PRP-3 provides the user with an end-to-end ARQ scheme where the source receives ACK/NACK responses for each packet sent in fixed-length strings.



(a) NORMAL TRAFFIC FLOW



(b) RETRANSMISSION DUE TO A NACK



(c) RETRANSMISSION DUE TO A TIME-OUT

IA 59-265

Figure 3-9. Packet Routing Protocol (PRP-3)

This protocol is best suited for error-sensitive data messages. PRP-3 offers the following advantages and/or features:

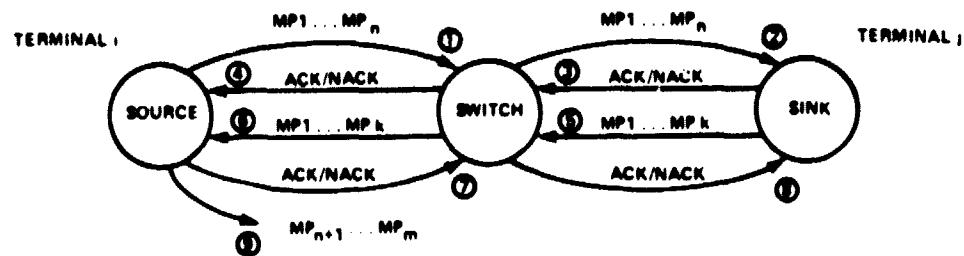
1. End-to-end ARQ protection is provided for each packet sent.
2. PRP-3 is more efficient for use with long messages than either PRP-1 or PRP-2.
3. Time-out hardware and/or software is not required on-board the satellite. The time-out system is only activated at the source for transmission sessions.
4. No error detection is required at the switch.
5. Packets are sent in user selectable packet string lengths.
6. Since ACK/NACK messages are used for packet strings, packet-size ACK/NACK messages can be used to avoid special size packets.

The disadvantages and/or overhead costs associated with PRP-3 are:

1. Buffering of all unacknowledged packets is required.
2. No uplink ARQ function is provided by the satellite.
3. PRP-3 is not needed for real-time voice messages since it supports retransmission of incorrect packets.
4. Time-outs require the retransmission of entire packet strings
5. Error-detection is required at the sink.
6. Error-correction is needed at the switch and the sink.

• PRP-4

PRP-4 is designed to handle voice traffic. The pattern of message traffic between a source and sink while under the control of PRP-4 is illustrated in figure 3-10. When the connection is established, the source begins sending his message, packet-by-packet. The sink sends the source a packet-length ACK/NACK message every time it either receives enough message packets to fill the ACK/NACK packet, or the session has terminated. The information contained in the ACK/NACK packet is used only for maintenance and



- ① THE SOURCE TRANSMITS A STRING OF PACKETS TO THE SINK.
- ② THE SINK RECEIVES THE STRING OF PACKETS
- ③ AN ACK/NACK IS SENT FOR EACH PACKET RECEIVED BY THE SINK.
- ④ THE SOURCE RECEIVES THE ACK/NACK MESSAGE. THIS INFORMATION IS USED FOR STATUS MONITORING AND BOOKKEEPING FUNCTIONS. SINCE THE PACKETS BEING ACKNOWLEDGED CONTAIN REAL TIME DATA, RETRANSMISSIONS MAY NOT BE POSSIBLE AND/OR MAY NOT BE NECESSARY.
- ⑤ SINCE THIS PROTOCOL SUPPORTS BIDIRECTIONAL COMMUNICATIONS, THE SINK IS ENTITLED TO TRANSMIT MESSAGES BACK TO THE SOURCE.
- ⑥ - ① THE SOURCE RECEIVES AND ACKNOWLEDGES THE SINK'S MESSAGE

8
39
12

Figure 3-10. Packet Routing Protocol (PRP-4)

channel status monitoring. This protocol can be used to support the transmission of real-time voice messages where retransmission is not possible and/or is not needed. Occasional bit errors or missing packets (<1% of total arriving packets if packets contain < 240 milliseconds of speech) (8) will not severely distort voice communications. Since most voice communications require two-way traffic, the source and sink change roles as the two parties communicate.

The benefits of PRP-4 are:

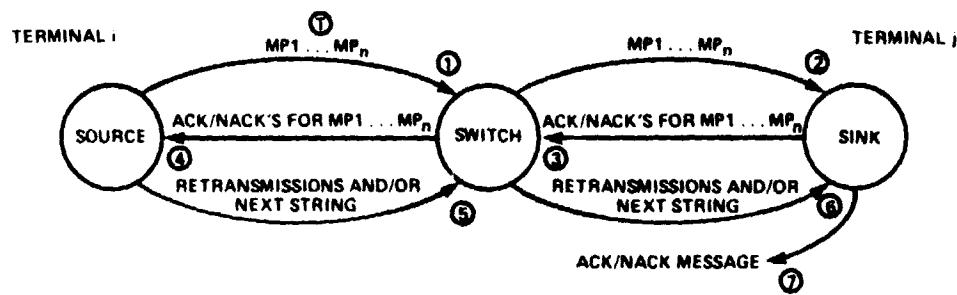
1. Real-time voice messages are supported by this protocol.
2. The source is not required to buffer packets once they are sent.
3. ACK/NACK information is still available to the network for status monitoring.
4. No time-out scheme is required.
5. No packet retransmissions are required.
6. Messages of any length can be supported.
7. Standard size packets can be used for the ACK/NACK messages.

The costs and disadvantages of PRP-4 are:

1. Error-sensitive messages are not supported.
2. Error detection is required at the source.
3. Error correction of the packet headers is required at the switch and at the sink.

• PRP-5

The flow of message packets as seen in figure 3-11 depicts the exchange of messages under the control of PRP-5. When the link is established, the source sends a string of packets to the sink. The length of the string is selected at link establishment time and remains fixed for the duration of the session. Each time a string is sent, the source starts a time-out clock and waits for the ACK/NACK message. After the last packet of the string is received, the sink sends the source a packet-length ACK/NACK message. All



- ① A STRING OF PACKETS IS SENT BY THE SOURCE TO THE SINK.
- ② THE SINK RECEIVES THE STRING.
- ③ AN ACK/NACK IS SENT FOR EACH PACKET IN THE STRING. THE SINK SENDS AN ACK FOR EVERY PACKET IT RECEIVES (REGARDLESS OF ERRORS). ONLY MISSING PACKETS RESULT IN THE TRANSMISSION OF NACKS.
- ④ THE SOURCE RECEIVES THE ACK/NACK MESSAGE.
- ⑤ THE SOURCE RETRANSMITS ALL PREVIOUSLY LOST PACKETS AND STARTS TO SEND THE NEXT STRING.
- ⑥ SINK BEGINS RECEIVING THE NEW STRING.

1859281

Figure 3-11. Packet Routing Protocol (PRP-5)

received packets are acknowledged whether or not they contain errors. Only missing packets are not acknowledged by the sink. The source retransmits any packets considered missing by the sink in the next string, along with new packets.

PRP-5 provides the users with an end-to-end ARQ scheme in which packets missing from a fixed-length string are retransmitted. This scheme is ideal for the transmission of facsimile or video data messages. Occasional bit errors can be ignored without severely degrading the image arriving at the sink. However, missing packets cannot be tolerated for this type of service.

PRP-5 provides the following services and benefits:

1. End-to-end ARQ protection is provided for each packet sent.
2. No error-detection is required.
3. Time-out clocks are needed only at the source.
4. PRP-5 is ideal for facsimile or video data messages.
5. Messages of any length can be supported.
6. Standard size packets can be used for the ACK/NACK messages.

The overhead costs and the disadvantages of PRP-5 are:

1. The sink must be able to detect missing packets.
2. All unacknowledged packets must be buffered.
3. Bit-error-sensitive messages are not supported.
4. Time-outs require the retransmission of the entire string considered lost.

PRP-1, PRP-2, and PRP-3 are suggested for error-sensitive data messages. PRP-4 is best suited for voice messages. Facsimile and video transmission are best supported by PRP-5.

The multiple protocol network provides the users with considerable flexibility. However, this flexibility implies additional hardware and software to support the various protocols. If some users do not send and receive all classes of message, they

may not want to support all the available PRPs in order to reduce the cost of their terminal or tariff.

3.4.3 Protocol Time-Out Parameters

Most of the protocols introduced in the previous sections rely on time-out clocks to notify the source of a message that the message has gone unacknowledged too long. In order to design the various time-out clocks, the amount of time they are to run before timing-out needs to be determined. Before these times can be calculated, the following parameters must be defined:

1. T_u = propagation time to the satellite (uplink)
2. T_d = propagation time from the satellite (downlink)
3. $T_{rt} = T_u + T_d$ = round-trip satellite propagation delay
4. T_{pc} = processing time at the controller (orderwire processor)
5. T_{pr} = processing time at the sink (receiver)
6. T_{pt} = processing time at the source (transmitter)
7. T_{ps} = processing time at the switch for a message packet or an ACK/NACK message packet
8. T_{ps1} = processing time at the switch for the single ACK/NACK packet (PRP-1 and PRP-2)
9. SM_i = safety margin i

An important point about the parameter definitions is that the processing times at each node include the demodulation times. The delays due to demodulation may vary depending on the packet lengths and the data rates.

3.4.3.1 The LEP Time-Out Values

The time-out values calculated in this section are based on the assumption that the controller is located on-board the satellite. If the controller were located in an earth station, these values would be larger as a result of extra propagation delays and additional processing times due to signal regeneration.

The LEP for point-to-point links requires two time-out clocks. One clock is needed for each query packet sent by the controller and the other time-out clock is used for each link request packet sent. Since the value of the link request time-out clock is a function of the query packet time-out, the value of the query time-out is determined first. The cycle of a query packet includes its transmission to the sink, processing at the sink, the transmission of a response packet back to the satellite, and the processing of the packet by the controller. Thus,

Time-out for a query packet =

$$Q_{To} \geq T_d + T_{pr} + T_u + T_{pc} + Sm_1 \quad (3.1)$$

The cycle of a link request packet includes the packet's transmission time to the controller, processing at the controller, waiting for the query/response cycle, the transmission of a link status packet back to the source, and the processing at the source. Thus,

Time-out for a link request packet =

$$LR_{To} \geq T_u + T_{pc} + nQ_{To} + T_{pt} + T_d + Sm_2 \quad (3.2)$$

where n = the number of allowed query packet time-outs.

The link LEP for multipoint links (broadcasting, conferencing) requires two time-out clocks. The clocks are needed for the special link status packet for partial multipoint links and for the multipoint link request packet.

The cycle of the special link status packet includes the packet transmission to the source terminal from the controller, the processing time at the source, and the transmission time of the link acceptance packet from the source to the controller. Thus,

Time-out for a link status packet for partial multipoint =

$$LS_{To} \geq T_d + T_{pt} + T_u + T_{pc} + Sm_3 \quad (3.3)$$

The cycle of the link request packet includes the transmission time of the packet from the source to the controller, the processing

time at the controller, the query/response cycle, and it may include the link status/link acceptance cycle. Thus,

Time-out for a multipoint link request packet =

$$LR_{To} \geq T_u + T_{pc} + n Q_{To} + m LS_{To} + T_d + T_{pt} + S_{m4} \quad (3.4)$$

where n = the number of allowed query packet time-outs and m = the number of link status time-outs.

3.4.3.2 The PRP Time-Out Values

The packet routing protocols require a time-out clock at the source of the message packets. This ensures that every source will receive either an ACK/NACK message or a time-out signal for each packet or packet string it sends.

The PRP-1 transmission cycle, which starts with the transmission of the message packet and ends with the source receiving an acknowledgment, entails quite a few stages. The message packet is first sent to the satellite, where it is processed and sent to the sink. At the sink, the packet is received and processed. The sink generates an acknowledgment message and sends it to the satellite. If the ACK/NACK message is error-free, it is then sent to the source terminal. However, if the ACK/NACK message contains errors, it is destroyed and the switch sends the sink a NACK. The sink is then required to retransmit the acknowledgment packet. Therefore,

Time-out for a message packet under PRP-1 control =

$$T_{PRP-1} \geq T_u + T_{ps} + T_d + T_{pr} + T_u + T_{ps1} + r(T_d + T_{pr} + T_u + T_{ps1}) + T_d + T_{pt} + S_{m5} \quad (3.5)$$

where r = the number of allowed uplink NACKs for a given ACK/NACK message. Equation (3.5) reduces to the following:

$$T_{PRP-1} \geq 2T_u + T_{ps} + 2T_d + T_{pr} + T_{ps1} + r(T_d + T_{pr} + T_u + T_{ps1}) + T_{dt} + S_{m5} \quad (3.5a)$$

Since $T_u + T_d = T_{rt}$, equation (3.5) can be further simplified,

$$T_{PRP-1} \geq 2T_{rt} + T_{ps} + T_{pr} + T_{ps1} + r(T_{rt} + T_{pr} + T_{ps1}) + T_{pt} + S_m 5 \quad (3.5b)$$

Since PRP-2 is PRP-1 without the uplink ARQ function, the time-out value is:

$$T_{PRP-2} \geq 2T_{rt} + T_{ps} + T_{pr} + T_{ps1} + T_{pt} + S_m 6 \quad (3.6)$$

PRP-3 operates much like PRP-2. The primary difference is that the ACK/NACK packets used in PRP-3 are the same size as the message packets. Therefore, all processing times at the switch are the same (T). An additional difference in the two protocols is that packets are sent and acknowledged in strings in PRP-3. However, since the time-out clock is not started until the last packet of a string has been sent, the equation used to determine the clock times is similar. Thus, the time-out clock value for the PRP-3 is:

$$T_{PRP-3} \geq 2T_{rt} + 2T_{ps} + T_{pr} + T_{pt} + S_m 7 \quad (3.7)$$

Since PRP-4 is used to support voice messages, retransmissions due to time-outs are not required. Therefore, no time-out mechanism is needed for this protocol.

PRP-5 operates much like PRP-3. Therefore, the time-out clock value for this protocol is:

$$T_{PRP-5} \geq 2T_{rt} + 2T_{ps} + T_{pr} + T_{pt} + S_m 8 \quad (3.8)$$

These equations can be used to determine the impact of packet delay times due to retransmissions. Since the network is based on a geostationary satellite, $T_u \approx T_d \approx 1/8$ second.

3.5 PACKET FORMATS

In order to support the two-level network protocol, two packet formats are needed. One type of packet is called the CONTROL

PACKET. These packets are used by the LEP to establish and to disengage circuits between terminals. There are many kinds of control packets. However, every control packet shares an identical format. This helps to keep the processing costs to a minimum by standardizing the system hardware and software.

The second type of packet is the MESSAGE PACKET. Every message packet consists of a HEADER, which contains the routing control signals and message reconstruction information, and a BODY. The body contains the actual message being sent.

3.5.1 Control Packet Format

All control packets, regardless of their function, share the same packet format. This helps to simplify processing requirements as well as hardware needs. Each control packet consists of four field categories. The various control fields within the control packet are displayed in figure 3-12.

The first category of fields is the CONTROL DATA category. The first field in this category is the PACKET TYPE field. This field identifies the type of control packet being used. The eight control packets associated with the LEP and their functions are:

1. LINK REQUEST - The link request packets are used by terminals that need satellite links to other terminals. The source terminal must send the link request to the controller to initiate the link establishment. The packet contains information regarding the message, the type of link desired, the source, the sink, the required routing protocol and the tariff charges. This packet is issued only by terminals which are to be the source of a call.
2. QUERY - The controller sends query packets to all terminals that have been specified as sinks in a link request. Each query packet contains the same information found in the link request that initiated the link establishment procedure. Thus, the sink is informed of the source's request and all the details concerning the pending message. In addition, upon receiving a query packet, a sink is obligated to inform the controller as to whether or not he is available to receive the message from the source. Query packets can be issued only by the controller.

CONTROL DATA			MESSAGE PARAMETERS					SOURCE/SINK				PACKET OVERHEAD	
CONTROL PACKET TYPE	STATUS CLOCK	SINK STATUS	MESSAGE MODE	MESSAGE TYPE	MESSAGE LENGTH	PACKET ROUTING PROTOCOL	PACKET SIZE	DATA RATE	SOURCE ADDRESS	NUMBER OF SINKS	SINK ADDRESS/ERROR DETECTION	PACKET ID	ERROR DETECTION
5 BITS	33 BITS	32 BITS	2 BITS	3 BITS	5 BITS	24 BITS	4 BITS	2 BITS	2 BITS	12 BITS	4 BITS	204 BITS	3 BITS

CONTROL DATA FIELDS
 MESSAGE PARAMETER FIELDS
 SOURCE SINK FIELDS
 PACKET OVERHEAD FIELDS
 PACKET LENGTH

Figure 3-12. Control Packet Format

3. RESPONSE - After receiving a query packet a sink sends the controller a response packet to indicate whether or not he is available to receive the source's message. This packet also contains the message and link parameter information provided by the link request. Response packets are generated by terminals that have been requested to be a message sink.
4. LINK STATUS - Once the controller receives a response packet from the queried sink, it sends the source a link status packet. Again, this packet contains a copy of all the message and link requirements specified in the original link request packet. The link status packet relays the sink's reply to the source. If the sink is busy, no link is provided. However, if the response packet indicates that the sink is available, the controller will establish the link and the source is free to transmit when he receives the link status packet. The source is not required to reply to a link status packet unless the link status indicates that the network is providing him with a partial multipoint link. When a response to a link status packet is needed, a special control packet is used. This packet is described next. Only the controller can send link status packets.
5. LINK ACCEPTANCE - The link acceptance packet is sent to the controller by a source who received a link status packet regarding a partial multipoint link. The link acceptance packet informs the controller as to whether or not the source wants the partial multipoint link. This packet is generated by terminals that are the source of a message.
6. TERMINATE CALL - When the source completes his message transmission, he sends a terminate call packet to the controller signalling the end of the call. The source terminal is the only node allowed to send this type of packet.
7. INTERRUPT - If network problems or special conditions arise that require the network to reclaim an allocated link, the controller sends the users an interrupt packet. This packet informs the users that they have been preempted and must stop transmission immediately. The source must send the controller a terminate call packet to indicate he has acknowledged the interrupt. The controller will send the users a link status packet as

soon as the network can re-establish a link. Once the users receive the link status packet, they may resume transmission. The network permits only the controller to issue this control packet.

8. HELLO - The hello packets are used by the controller for testing and monitoring the status of the network. The controller can send a hello packet to any user in the network at anytime. When a user receives a hello packet he must send a hello packet back to the controller as soon as possible. Using hello packets, the controller can determine delay times, user availability, channel quality, and the location of link/node failures.

The next field in the control data category is the STATUS/CLOCK field. This field allows the users and the controller to send either status information or clock synchronization data to one another. The first bit in the field identifies the next 32 bits as either status or clock data. When this field is used in a hello packet, it contains clock data so that the controller can measure delay times.

Following this field is the SINK STATUS field. This field is organized as 16 two-bit words. The need for the 16 status words stems from the fact that in multipoint communications, a source may request up to 16 sinks. When a sink replies to a query packet he must update the sink status word which corresponds to the position of his address in the sink address field (this field is explained later). For example, if a sink's address is the third address in the sink address field, he must update the third status word in the sink status field. Each word can assume one of four states. State 1 is found in all link request and query packets. States 2 through 4 are updated by the queried sinks and are found in all response and link status packets. States 3 and 4 are found in all Hello packets sent to the Controller by a user. When a user replies to a Hello packet sent by the controller, no message follows. The Hello packet is used only for status monitoring. This packet does not initiate a link establishment. The four states and their definitions are:

1. NO RESPONSE - This indicates that the sink has not responded to the query yet. This state should only be found in link request and query packets since it is an initial state.
2. UNACCEPTABLE CONDITIONS - This message is used to inform the source and the controller that the sink may be willing to accept the message if certain transmission parameters are changed. When the sink issues this control signal, he

must indicate the parameters he wants changed. He does this by replacing the current parameters with the ones he would like to operate under. If the source is able to meet the new requirements, he sends the controller a new link request packet (which now specifies the new conditions requested by the sink) and the link establishment procedure begins again. Otherwise, the link is not established and the source must decide what to do about its incompatibility with the desired sink.

3. SINK BUSY - When the sink is busy with another call or if he is unable to receive a message due to other reasons, he selects this message to send in a response packet or hello packet.
4. SINK READY - A sink generates this message when he is available and ready to receive the pending message. A user also sends this message in the hello packet if he is currently available to receive a message if one did exist.

3.5.1.2 Message Parameter Fields

The next group of fields found in the control packet is MESSAGE PARAMETERS. The fields in this category allow the users to specify a message and how it is to be sent. This information is vital to the controller as well as the sink. The controller uses the information to establish the link while the sink requires the information to prepare for pending transmission. The first field in this category is MESSAGE PRIORITY. This allows the users to assign one of four priorities to their messages. The higher the level of priority a message has, the lower the delay times the packets of that message experience. These options are offered to the users at various tariffs. The four levels of priority are:

1. No priority (standard tariff)
2. Low Priority
3. Priority
4. High Priority

The second field is the MESSAGE MODE. Users can select one of three modes that use either a SIMPLEX, HALF-DUPLEX or a DUPLEX circuit. A simplex circuit provides one-way transmission while the duplex circuit provides simultaneous two-way communications. The half-duplex circuit provides two-way communications by providing

message transmission in one direction at a time. The three message modes used in conjunction with the selectable circuits are:

1. POINT-TO-POINT - This message mode connects a single source to a single sink. The role of source and sink may interchange during a call if a duplex or half-duplex circuit is used.
2. BROADCASTING - Broadcasting allows a user to send a message to as many as 16 sinks. This mode does not require duplex circuits since it is considered as a "one-to-many transmission". (The sinks may send the source ACK/NACK messages if they are required by the packet routing protocol being used.)
3. CONFERENCING - Duplex circuits are required in this mode since any one of the up to 16 sinks may send a message as well as the original source. (Half-duplex circuits may also be used, but this may require additional hardware/software complexity.) In this mode, every connected user has the ability to transmit a message and to receive all the messages sent by the other users during the duration of the call.

MESSAGE TYPE is the third field found in the message parameters category. A user can specify one of 32 message types. The more well known message types are voice, data, video, and facsimile. Each of these message types can be specified in more detail by the user. The user can do so by selecting one of the message prefixes provided in this field such as real-time, non-real-time, store-and-forward, and batch. In addition, various tariff rates could be selected as well. Thirty-two options should provide ample flexibility for the users to vividly specify any message they will ever need to send.

Next to the message type field is the MESSAGE LENGTH field. This field allows the source to specify any message length up to 1.49×10^9 bits. The message length is encoded using two blocks within the field that total 24 bits. The first block is 4 bits wide and is used to encode the units of the message. A user can select one of 11 message units. The units available are:

1. Bits
2. Nibbles (4 bits)
3. Byte (8 bits)

4. Words (16 bits)
5. Double words (32 bits)
6. 64 bits
7. 128 bits
8. 256 bits
9. 512 bits
10. 1024 bits
11. 1424 bits (one full packet at T1 rates)

After a user selects the message unit, he determines the number of units in the message and he encodes this in the second block, which is 20 bits wide. (when selecting a message unit, a user must keep in mind that an entire message can never exceed 2^{20} units.) By using this two level scheme, a 24-bit wide field can now specify messages that are slightly longer than 230 bits in length rather than 2^{24} when straight decoding is used.

The PACKET ROUTING PROTOCOL field is provided to the users so they can select the PRP best suited for the particular message pending transmission.

The PACKET SIZE field contains the size of the packet to be used for transmission. This parameter can be changed by the user on a message-by-message basis. However, the size of a packet must remain constant throughout a message session once it has started.

The last field in the message parameter block is the DATA RATE field. Users are expected to transmit at the standard T1 rate of 1.544 Mb/s unless they notify the network via this parameter field.

3.5.1.3 Source/Sink Fields

The source/sink fields are used to identify the source and sink(s) of a message. The first field is the SOURCE ADDRESS field. This field contains a 12-bit encoded address of the user who initiated the LEP. This field does not identify the source of the control packet itself, in every instance. The controller can be the source of four types of control packets, but its address is placed in the source address field only for hello packets. Similarly, the sinks can be the source of two types of control packets, but their addresses are found in the source address field only for hello

packets. Since each unique LEP control packet can only originate from one type of node (either a source terminal, a sink or the controller) the network can always determine the type of node that generated the control packet by knowing what type of packet it is. The second field in the source/sink block is the NUMBER OF SINKS FIELD. In this field the source indicates the desired number of sinks (within a range of 1 to 16). The next field, the SINK ADDRESS field provides the source with 16 slots (12 bits/slot) in which to place the address of each requested sink. When a sink sends a control packet, it clears this field and places its own address in the first slot. This permits the controller to determine who is the source of the Response packet it has just received.

The sink address field has a 17th slot that is reserved for the controller. When a source specifies that he wants a multipoint circuit, the controller issues a special 12-bit MULTIPONT ADDRESS CODE (MAC). The controller places this MAC in the query packets and in the link status packets thereby notifying the source and each sink of the special address they must use for the duration of the multipoint call. This single 12-bit address replaces the individual addresses of all the specified sinks. Thus, when the source begins transmitting the actual message, he addresses each message packet with the MAC rather than the actual sink addresses. When the switch sees the special code, it routes the packets to the associated sinks. Each sink linked to the multipoint circuit accepts all incoming packets containing the proper MAC rather than their own addresses for the duration of the call.

Almost 90% of the messages in the network are expected to be sent using the point-to-point mode which requires only one sink address. Thus, 90% of the time the remaining 16 slots can be used to send special control signals, and/or be used to send extra error detection bits (improving the probability of detecting multiple bit errors correctly).

3.5.1.4 Packet Overhead Fields

The last block of fields in the control packet is for PACKET OVERHEAD. These fields are required to maintain the integrity of the control packets. The first field is the PACKET ID field which is used to tag each packet with an ID number. This enables the network to keep track of the packets. The last field is the ERROR DETECTION field, which contains the bits required as overhead for an error-detection code. The code bits in this field enable the network to determine the validity of a received control packet.

All control packets are 380 bits in length. The network collects the users' data over a 1 millisecond frame. If the users

are transmitting at the T1 rate of 1.544 Mb/s, 1544 bits are collected in one frame. Thus, at the T1 rate packets are 1544 bits long. Therefore, the 380 bits must be made to fill the 1544 bit packet. Pseudorandomly spreading the control packet by a factor of four yields 1520 bits. This provides 24 bits which can serve as a preamble for each packet.

An important point to note is that the burst rate used on the satellite link is not T1. Each user's packet is burst up to the satellite within a 125 microsecond slot. Using the one millisecond frame, this slot allows the network to support eight users on one channel. 10 microseconds of the slot are used as an interpacket gap leaving 115 microseconds for the actual packet transmission. Thus, given the 125 microsecond dwell time with its 10 microsecond interpacket gap and the packet size of 1544 bits, the satellite link uses a burst rate of 13.4 Mb/s per channel.

3.5.2 Message Packet Format

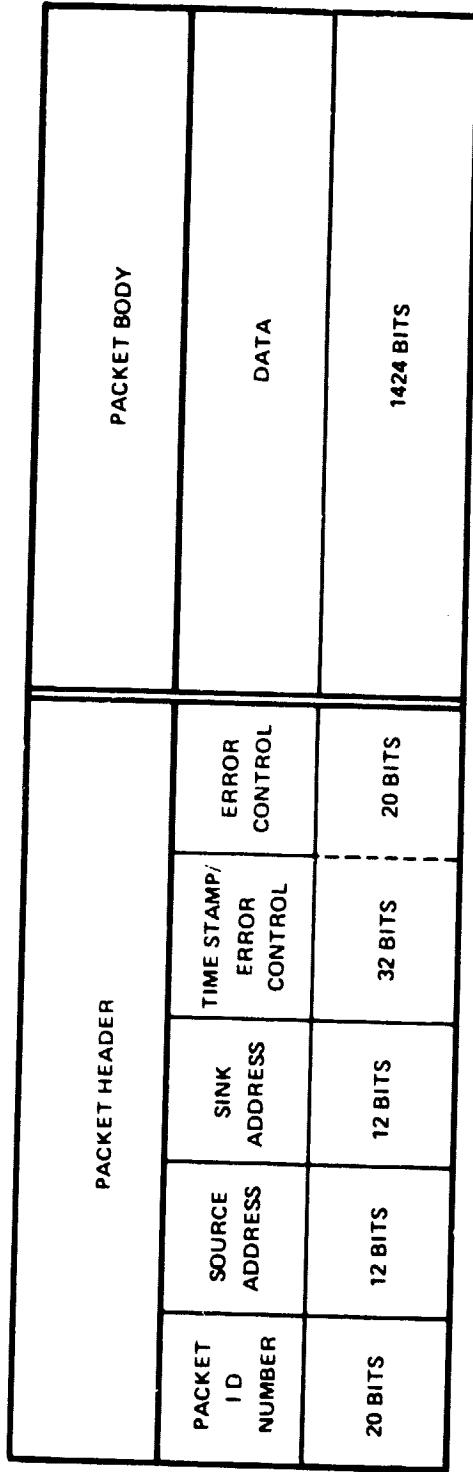
Each message packet is composed of a header and a body. Regardless of the body size chosen by the users, the header size remains constant. This eliminates the need for any special hardware or software to handle the headers. Since the header is fixed in length for every packet, the length of the body depends on the chosen data rate and satellite beam dwell time. Figure 3-13 contains the message packet fields and their bit requirements. A description of the fields follows.

3.5.2.1 The Header Fields

The first field in the header is the PACKET ID NUMBER field. Since a user may specify a message that can have as many as 2²⁰ units (where the maximum unit length is equal to the largest packet body allowable for the given frame time and user data rate), at least 20 bits are required if each packet in the message is to have a unique ID number. This field is essential since packets may arrive out of sequence, may not arrive at all, or may arrive with errors. Without this field, sinks would be unable to reconstruct their received message from the scrambled packets, or to request the retransmission of missing or corrupted packets.

The second field found in the header is the SOURCE ADDRESS field. This field contains the 12-bit encoded address of the user sending the packet. A sink requires this address whenever it must ask for a retransmission. The network could use this information for control, routing, and/or billing.

IA 59.264



PREAMBLE = 24 BITS
HEADER = 96 BITS
BODY = 1424 BITS

PACKET BLOCK = 1544 BITS

Figure 3-13. Message Packet Format

The third field is the SINK ADDRESS field. This field consists of the 12-bit address of the sink. The network uses this address to route the packet to its final destination. In the case of multipoint calls, the multipoint address code (MAC) is placed in this field. All the users of the multipoint circuit are aware of this code and will accept messages addressed with that code. The network issues the MAC and is responsible for the proper routing of the packets to their multiple destinations.

The fourth header field is the TIME STAMP field. This field provides a 32-bit time stamp for the packet. Packets carrying voice information are time stamped so that when they reach the sink they can be sorted. In addition to being used for sorting, the time stamp is compared to the current clock value at the sink. If the time difference exceeds a predetermined limit, the packet is discarded. When a time stamp is not required by a PRP, this field is used in conjunction with the fifth field which is the ERROR CONTROL field. The 32-bit time stamp field is concatenated with the 20-bit error control field to provide a 52-bit error-correction code. When a time stamp is necessary, only the error control field is used to hold the error-correction bits. The error control field is the last field in the packet header bringing the total length to 96 bits.

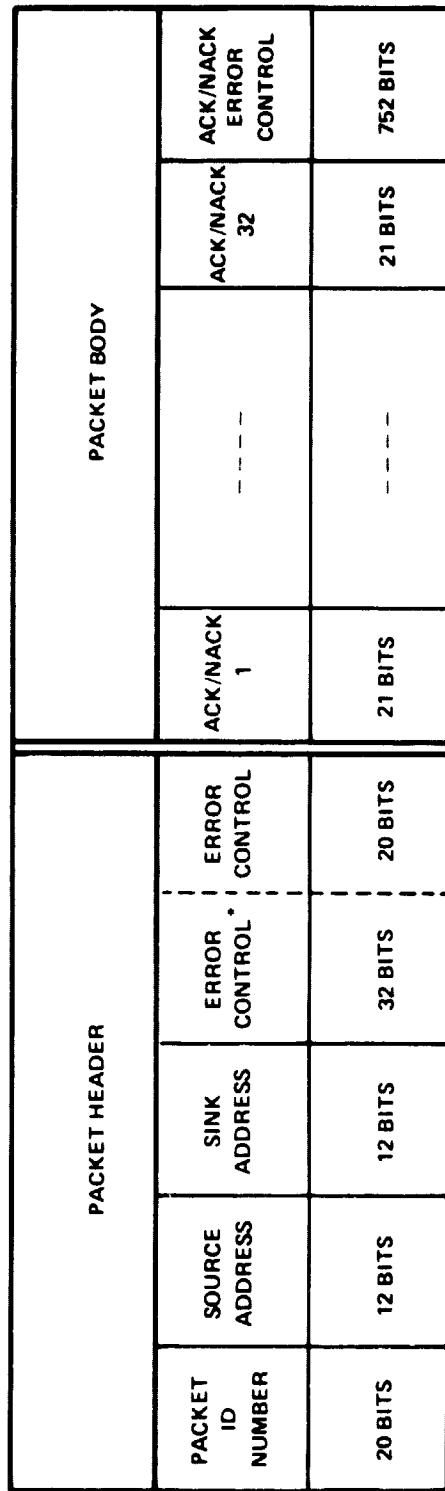
3.5.2.2 The Packet Body

The body of the message packet consists of the DATA FIELD. The data field is reserved to carry the actual text of a user's message. The user is entitled to use this field to send any type of message that can be supported by an available PRP.

Like the control packets, message packets are preceded by a 24-bit preamble and are received by the network in 1 millisecond frames. Thus, if the users are transmitting at the T1 rate of 1.544 Mb/s, the total packet length is 1544 bits. Since the preamble is 24 bits long and the header is 96 bits wide, the packet body must be 1424 bits in length. Each packet is bursted up to the satellite in 125 microsecond slots (10 microseconds of guard time and 115 microseconds transmission time) resulting in a satellite uplink burst rate of 13.4 Mb/s.

3.5.3 ACK/NACK Message Packet Formats

Figure 3-14 displays the packet format for an ACK/NACK message packet. The ACK/NACK message packet is based on the message packet format. Therefore, although the ACK/NACK message packet carries out a different function, it is considered a type of message packet.



*ERROR CONTROL OPTION FOR THE TIME STAMP/ERROR CONTROL FIELD IS ALWAYS SELECTED.

PREAMBLE	=	24 BITS
HEADER	=	96 BITS
BODY	=	1424 BITS
<hr/>		
PACKET BLOCK	=	1544 BITS

Figure 3-14. ACK/NACK Message Packet Format

This helps reduce processing costs since all packets traveling through the network share the same header format (and the same length when the data rate is T1 on the uplink in 1 millisecond frames.) As seen in figure 3-12, the first three fields in the header are identical to those in the message packet. The major difference in the ACK/NACK header is that the fourth field, ERROR CONTROL, always uses the 32-bits of the time stamp field as additional error-correction bits, bringing the total to 52 bits.

The primary difference in the ACK/NACK body is that the body is now divided into two fields rather than the single data field of a message packet. The first field, the ACK/NACK field, contains 32 blocks. Each block is further divided into two slots. The first slot is a single bit wide. The second slot contains 20 bits. The first bit serves as the actual ACK/NACK word while the remaining 20 bits contain the ID number of the corresponding message packet. Thus, using all the ACK/NACK blocks, a sink may send the ACK/NACK responses for 32 received packets back to the source in a single ACK/NACK message packet. This field is 672 bits wide.

Assuming the sink is using the standard packet length of 1520 bits (preceded by a 24-bit preamble), the second field of the body must be 752 bits wide. This large field, the ACK/NACK ERROR CONTROL field contains the error-correction bits which protect the 32 ACK/NACK messages. At first glance, the fact that more than half of the bits in an ACK/NACK message packet are used for error control may seem wasteful. However, uncorrected errors in an ACK/NACK message can result in either the retransmission of the wrong message packets or the destruction, by the source, of a packet that the sink never received correctly. These problems contribute to the inefficient use of network resources as well as to the degradation of system performance and reliability. By using additional overhead in control messages such as ACK/NACKs, system performance can be enhanced while network resources are used efficiently.

3.5.4 Short ACK/NACK Packet Formats

As pointed out in section 3.4.2, some PRPs require an ACK/NACK reply for each message packet received. This constraint makes the use of full size packets for ACK/NACK response impractical. The need for a smaller packet arises when these PRPs are used. Figure 3-15 shows the smaller packet format used for single ACK/NACK replies. The entire ACK/NACK packet is the size of a standard message packet header (96 bits).

The first field contains the 20-bit packet ID number for the ACK/NACK packet itself. The second field is known as the SOURCE

IA-59,268

ACK/NACK PACKET ID NUMBER	SOURCE ADDRESS	SINK ADDRESS	ERROR CONTROL	ACK/ NACK	MESSAGE PACKET ID NUMBER
20 BITS	12 BITS	12 BITS	31 BITS	1 BIT	20 BITS

PACKET LENGTH = 96 BITS.

Figure 3-15. ACK/NACK Packet Format

field. This field contains the 12-bit address of the terminal sending the ACK/NACK. The SINK field, next to the source field, contains the 12-bit address of the user who is to receive the ACK/NACK reply. The fourth field is the ERROR CONTROL field. This field holds 31 error-correction bits which are used to protect the packet's information bits. If there is no need for the users to tag their ACK/NACK packets, they can use the 20-bit packet ID field as an additional error control field. This provides the users with 51 error-correction bits. The last field is the ACK/NACK field. This field consists of two blocks. The first block is one bit wide while the second block contains 20 bits. The single bit is the ACK/NACK word. The 20 bits found in the second block form the packet ID number of the received packet for which the reply is being sent.

3.6 SIZING THE BASEBAND PROCESSOR

The baseband processor services 16 uplink and 16 downlink beams. There are eight fixed beams and eight scanning beams for both the uplink and the downlink. The fixed beams can be implemented by using scanning beams that are programmed to scan the same coverage area continuously. All transmissions over the satellite links are synchronized within 1 millisecond frames. Each frame consists of eight scanning beam dwell intervals (slots). Every slot is 125 microseconds in length. Ten microseconds of each slot is reserved as an interblock gap (message guard band) while the remaining 115 microseconds are used for the actual transmission.

The satellite links are organized so that each uplink beam contains eight FDMA channels per slot. Each downlink beam contains eight TDM channels per slot. A channel is defined as an instantaneous signal from or to a distinct user or terminal, from the satellite's point of view. Once assigned to a particular slot, a channel must share its slot with seven other channels.

On board the satellite, there is one demodulator for each uplink channel. The downlinks have one TDM modulator per beam. Since there are 16 uplink and 16 downlink beams that each support eight channels/slot, signals on 128 unique channels will be arriving at the satellite simultaneously while 128 downlink channels are also active during each slot. Therefore, the baseband processor must contain 128 demodulators and 16 modulators.

By using baseline assumptions concerning the beams, the throughput and switching requirements of the processor can be determined. The following assumptions are made concerning the satellite beams:

1. 16 uplink and 16 downlink beams to be serviced by the baseband processor
2. frame period = 1 millisecond
3. 8 slots per frame => 1 slot = 125 microseconds
4. 8 channels per slot
5. channel capacity = 1 packet per frame

The uplink throughput of the baseband processor is:

Total number of packets per frame =

$$\text{No of beams} \times \left[\left(\frac{\text{No of channels}}{\text{slot}} \right) \times \left(\frac{\text{No of slots}}{\text{frame}} \right) \times \left(\frac{1 \text{ packet}}{\text{channel}} \right) \right] \text{ / beam (3-9)}$$

$$= (16) \times (8) \times (8) \times (1) \text{ packets/frame}$$

$$= 1024 \text{ packets/frame.}$$

If we allocate one packet user, per frame 1024 users can be serviced by the baseband processor:

Total number of users serviced =

$$\left(\frac{1}{\text{No of packets}} \right) \times \left(\frac{\text{No of packets}}{\text{frame}} \right) = \frac{1}{1} \cdot 1024 \text{ users} \quad (3-10)$$

An additional baseline assumption is that users are to transmit at an average T1 rate of 1.544 Mb/s. Using this value, the requirements of the satellite links and the baseband processor can now be specified in more detail. Since a user's data is collected over a 1 millisecond frame interval, the packet size for the T1 data rate is 1544 bits. This packet is sent over the satellite link within one slot. Since the uplink slots provide the networks with a 115 microsecond transmission period, a user's channel must be capable of an uplink burst rate of 13.4 Mb/s. Each uplink beam supports eight channels. Therefore, the total uplink bandwidth per beam is 107.4 MHz if a 1 b/s/Hz modulation is employed. Since each downlink beam supports eight TDM channels at eight times the uplink burst rate, the downlink bandwidth per beam is also 107.4 MHz, again

assuming a 1 b/cycle modulation. Since the processor services 1024 T1 users, the satellite throughput is $1024 \times 1.544 \text{ Mb/s} = 1.6 \text{ Gb/s}$.

3.7 THE PROCESSOR ARCHITECTURE

Figure 3-16 contains a simplified block diagram of the baseband processor. Each of the 16 on-board RECEIVERS is assigned to one of the 16 uplink beams. Each receiver separates and demodulates the eight FDMA channels arriving within the beam for each slot. The serial bit streams from each channel pass through an INPUT SWITCHING MATRIX. Each matrix handles four channels. The purpose of the matrix is to ensure that every channel can be connected to a SWITCHING PROCESSOR. The paths through each matrix are established by the on-board SYSTEM CONTROL COMPUTER. Generally, once these paths are established they remain fixed. However, if the need arises due to hardware failures, uneven traffic distributions, or related network problems, these paths can be altered on command from the Master Control Center. Table 3-3 contains all the possible interconnections provided by the switching matrices.

Since every switching processor provides service to one uplink channel and one downlink channel, the baseband processor employs 128 active processors. Thirty-two additional switching processors are available to the baseband processor. These processors provide a 25% redundancy, bringing the total to 160 switching processors. These switching processors are implemented in the form of microprocessors which each require local read only memory (ROM), random access memory (RAM) and external support hardware. The switching processors receive special control signals from the system control computer. At the on-set of each new message the ORDERWIRE PROCESSOR (CONTROLLER) provides protocol information, message length data, routing procedures, and special handling instructions to the proper switching processors. The controller and the system control computer communicate with one another in order to maintain synchronization. When both units are located on the satellite they are known as SYSTEM CONTROL (SYSCON).

The switching processors have four principal functions. One function is to store arriving packets in available locations in the satellite's BULK MEMORY (BM). A BULK MEMORY CONTROLLER (BMC) maintains a list of empty locations, full locations, and locations out of service. The switching processors send the BMC the addresses of locations no longer in use. In addition, when needed, the switching processors are issued addresses of available locations from the BMC. While an incoming packet is being stored into an available bulk memory location, the switching processor begins its second function. This function consists of processing the packet's header.

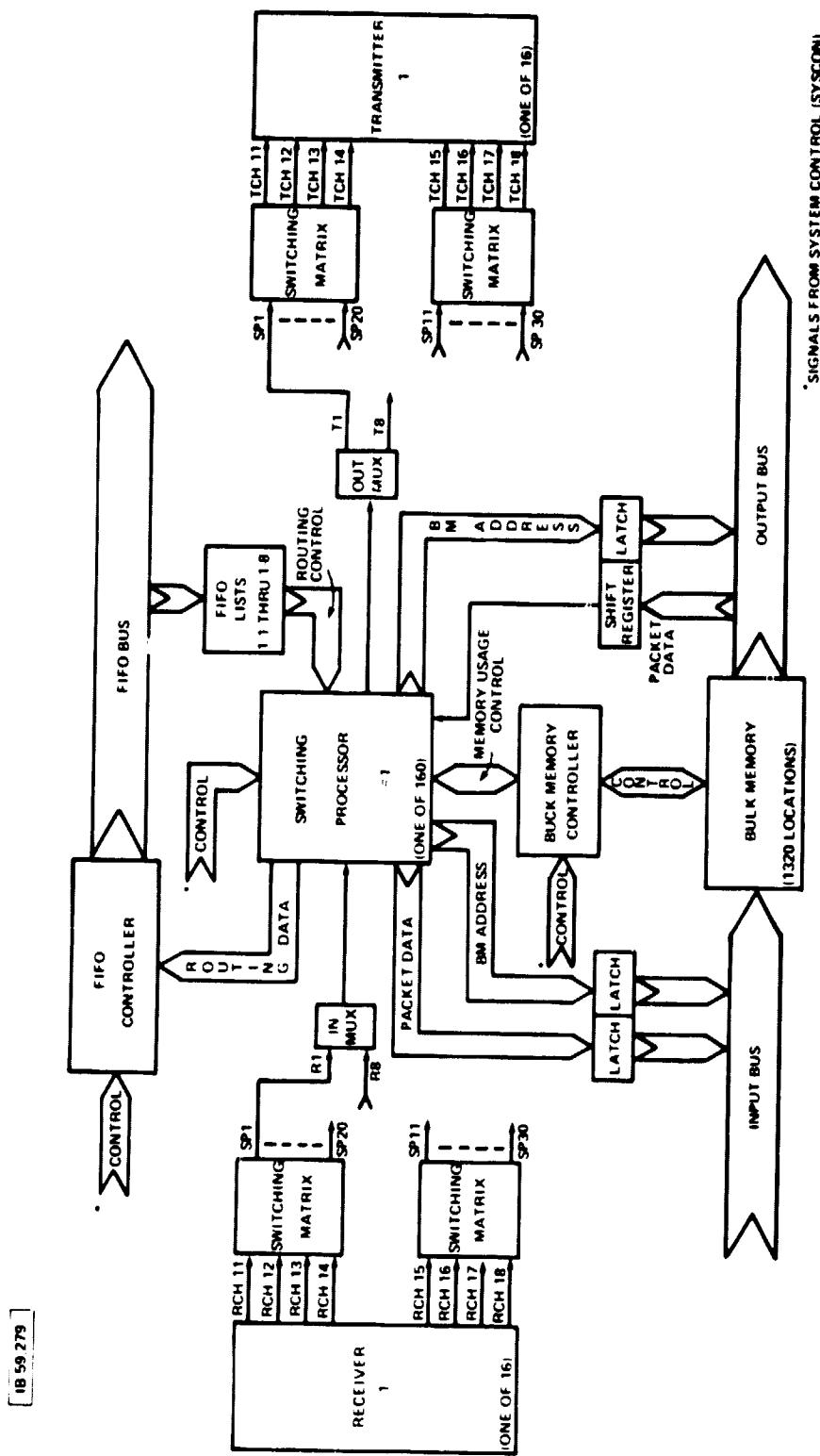


Figure 3.16. SIMPLIFIED BLOCK DIAGRAM OF THE BASEBAND PROCESSOR

Figure 3-16. Simplified Block Diagram of the Baseband Processor

Table 3-3

**Connectivity for the Receiver/Transmitter
Channels to the Switching Processors.**

Any of These Receiver/Transmitter Channels	Can Be Served By Any Of These Switching Microprocessors
11 - 14	P1 - P20
15 - 18	P11 - P30
21 - 24	P21 - P40
25 - 28	P31 - P50
31 - 34	P41 - P60
35 - 38	P51 - P70
41 - 44	P61 - P80
45 - 48	P71 - P90
51 - 54	P81 - P100
55 - 58	P91 - P110
61 - 64	P101 - P120
65 - 68	P111 - P130
71 - 74	P121 - P140
75 - 78	P131 - P150
81 - 84	P141 - P160
85 - 88	P151 - P160 + P1-P10

During the processing the switching processor corrects the header, determines the packet's destination, determines if the packet has multiple destinations and carries out PRP-related functions. After determining the packet's destination, the switching processor sends this information to the FIRST-IN-FIRST-OUT (FIFO) CONTROLLER. The FIFO controller (FFC) routes each packet by placing the packet's bulk memory addresses into FIFO lists. Each microprocessor is responsible for eight FIFO lists (one for a specific channel in each of the eight downlink slots). The switching processor's fourth function starts when the processors access the FIFO corresponding to the current slot. If the FIFO contains data, the processor fetches the oldest address in the FIFO. Using this address, the processor fetches a packet from bulk memory and sends it to one of the TRANSMITTERS where it is beamed down to the proper sink.

3.8 PROCESSOR OPERATION

In order for a user to use the baseband processor, he must first communicate with the controller via the LEP (see section 3.4.1). Once in contact with the controller, the user requests a satellite link to one or more sinks. The controller checks with the system control computer to determine if the network has the required resources available for the user's link. If the resources are available, the controller then checks with the desired sink(s). If the sink(s) is free, the controller establishes the message link. The link is completed when controller informs the system control computer, the user, and the appropriate switching processor.

After the user is allocated the message link, he formats his message into packets and begins transmission. The operation of the baseband processor is explained by tracing the path of one of this user's packets through the processor.

3.8.1 Point-To-Point Transmissions

After receiving a link status packet that indicates the satellite link has been established, the user prepares for transmission. Meanwhile, the switching processor assigned to handle the user's call is also preparing for the call. The processor carries out this preparation during the 10 microsecond interblock gap of the slot belonging to the user. During the interblock gap the processor fetches the address of an available BM location from the BMC. The processor then places this address into the Input Address Register in the BM array.

As the packet arrives at the demodulator it is converted into a bit stream. The packet enters the baseband processor serially through an input switching matrix. This matrix connects the output of the demodulator to the proper switching processor. At the switching processor, the serial bit stream is collected in a serial-to-parallel shift register. When this shift register is full, the 32-bit word is transferred to the Data Input Register. This is a 32-bit parallel-in-parallel-out register. Both the Input Address Register and the Data Input Register are connected to the bulk memory input address/data bus. A hardware poller in the bulk memory array sequentially enables the contents of each switching processor's register pair onto the bus. During this polling cycle, the address of the bulk memory is gated onto the bus from the enabled input address register. The 32-bit word of the packet is also strobed onto the bus from the selected input data register. The BM address decoder decodes the BM address currently on the bus and enables one bulk memory data latch. This latch fetches the 32-bit word from the bus and holds it until the slower bulk memory device can fetch the data from the latch. The hardware poller scans the registers of every switching processor. 32 bits are collected from the demodulators every 2.5 microseconds. Thus, the hardware poller must complete each cycle once every 2.5 microseconds before the new data over-writes the old data. At the end of the entire 125 microsecond slot, every received packet has been stored in bulk memory.

While each 32-bit word is being stored, the switching processor makes a copy of the packet's header and stores it in local RAM. After the entire header has been duplicated and stored into local RAM, the switching processor processes the header. This processing includes all PRP-related functions, error correction, and the decoding of the packet's destination address. After the sink address is decoded, the switching processor sends it and the packet's bulk memory address to a buffer in the FFC. As soon as the buffer is loaded, the processor sets a flag to indicate to the FFC polling circuit that the buffer contains valid routing information.

The FFC is responsible for the proper routing of every received packet. A hardware poller scans the buffers of each active switching processor. The contents of the enabled buffer are gated onto the FIFO controller bus. The address of the packet's destination is sent to a RAM containing a routing table as an address for the RAM. The contents of the RAM location addressed provides the address of one of the FIFO lists. Since this translation is done in a RAM, routing updates can easily be made by SYSCON. The FIFO address is placed onto the FIFO bus which is connected to every FIFO list. The address of the BM location of the packet is sent directly from the FFC buffer to the FIFO bus.

When a list decodes its own address from the FIFO bus, it latches the BM address from the bus. Once a packet's BM address resides in a FIFO list, it is considered routed.

Every switching processor has eight FIFOs. Each FIFO corresponds to one unique channel in a unique downlink slot. During the 10 microsecond interblock gaps, the processors check the FIFO lists corresponding to their channel in the next slot. If a processor finds that the FIFO list is not empty, the processor fetches the oldest BM address from the list. This address corresponds to the location in bulk memory which contains the oldest packet destined to the sink being serviced by the channel in the next slot. This address is placed in the Output Address Register by the processor. Working in conjunction with this register is a 32-bit wide parallel-in-serial-out register called the Output Data Register.

At the start of the transmission cycle, a hardware poller in bulk memory scans each Output Address Register. Each time the address register is polled it gates its contents (a BM address) onto the bulk memory output address/data bus. This address is supplied to each memory module. If a memory module recognizes its address on the bus, it sends the next 32-bit word of the packet it is storing to the enabled Output Data Register via the bus. (Each time an Output Address Register is polled, its corresponding Output Data Register is also enabled by the poller.) The enabled register then latches the data and begins shifting it out serially.

The serial bit stream is routed through the processor's output multiplexer, through an output switching matrix, and into a 1544-bit (packet-length) serial-in-serial-out shift register. This register feeds a modulator at the rate of about 100 Mb/s so that the packet can be bursted down to its sink on the selected channel within the proper slot on a TDM basis. Each switching processor is actually responsible for two of the serial-in-serial-out shift registers. One is being filled in advance by the processor while the other is being emptied by the remodulator.

The polling cycle in the bulk memory occurs once every 2.5 microseconds until each addressed packet is transferred out of the bulk memory 32-bits at a time and sent to the proper remodulators for transmission. This packet transfer requires forty-nine 32-bit word transfers and a total of 125 microseconds to complete. Therefore, in order for a packet to be transmitted down on the desired channel during the proper slot, it must be sent to the remodulator exactly one slot prior to its transmission.

3.8.2 Multipoint Transmissions

After receiving a link status packet that indicates that the satellite links have been established, the user prepares for transmission. Meanwhile, the switching processor assigned to handle this call is also preparing for the message transmission. During the user's interblock gap, the processor fetches the address of an available bulk memory location from the BMC. This memory location must be capable of handling packets that are destined for multiple sinks. The processor then places this address into its input address register. The switching processor then sends the same address back to the BMC along with the number of sinks that the packet is destined for. The number of sinks is supplied to the processor by SYSCON at the onset of the link establishment. The BMC needs this information to ensure that no memory location is reused until all the specified sinks have received a copy of the multipoint packet.

As a packet arrives from its demodulator, it is sent into the bulk memory array by the switching processor. While each 32-bit word is being collected and routed to the proper BM location, the switching processor makes a copy of the packet's header and stores it in local RAM. After the entire header has been duplicated the processor begins decoding and processing the header. The processor determines the packet's MAC and sends it to the buffer in the FFC along with the packet's BM address. Also, the multipoint alert flag is set by the processor.

A hardware poller in the FFC scans each of the processor's input FFC buffers. The BM address from an enabled buffer is sent directly to the FIFO bus. The MAC is routed to the MAC decoder due to the set multipoint alert flag. The MAC decoder contains a list of each FIFO which is to be updated with the BM address currently on the FIFO bus. This list is supplied to the FFC by SYSCON during the link establishment procedure. Each FIFO address in the list is gated onto the FIFO bus one at a time for a fixed duration. The BM address remains stable on the bus throughout the entire multiple FIFO update. As each FIFO address appears on the bus, the corresponding FIFO fetches the BM address and places it into the list. This scheme permits the network to provide multipoint addresses using a single copy of the packet and a single sink address in the packet header.

During each interblock gap, every active switching processor checks its own FIFO list corresponding to its channel in the next slot. If there is a packet awaiting transmission, its BM address resides in the FIFO list associated with the packet's destination. After fetching the BM address of a packet, the switching processors

send it to their output address register in the BM array. The processors then access the BM modules and begin sending the packets to the modulators being served by the processors. Multipoint packets may require "simultaneous" (within the same slot) BM accesses by several switching processors. Therefore, special BM modules, which contain additional control hardware, are allocated for multipoint packets in order to solve the contention problems. These locations are capable of supplying more than one switching processor with a copy of the same packet within the same slot.

After routing the packet from memory to the proper modulators, the switching processors return the BM address to the BMC. The BMC counts the number of times the same BM address is returned. The BMC compares this value with the number of sinks that are to receive a copy of the packet. A multipoint BM location is not released by the BMC until a copy of the packet at that location has been sent to every specified sink.

3.8.3 Implementation of the Baseband Processor

The architecture and the chip-level implementation of the baseband processor/packet switch presented in this report represent one possible system design. This design may or may not be an optimal solution. However, using this design as a guideline, power consumption can be estimated, an approximate chip count can be made, and most importantly, technology requirements for future on-board satellite processing capabilities can be identified.

3.8.3.1 The Bulk Memory Controller

The BMC has three major tasks:

1. To monitor the status of every bulk memory location.
2. To accept the addresses of unneeded BM locations from the switching processors.
3. To provide the addresses of available BM locations to the switching processors on demand.

Figure 3-17 contains a block diagram of the BMC. All switching processors are required to return the address of a BM location they have been accessing whenever:

1. They are not scheduled to receive an incoming packet in the next slot.

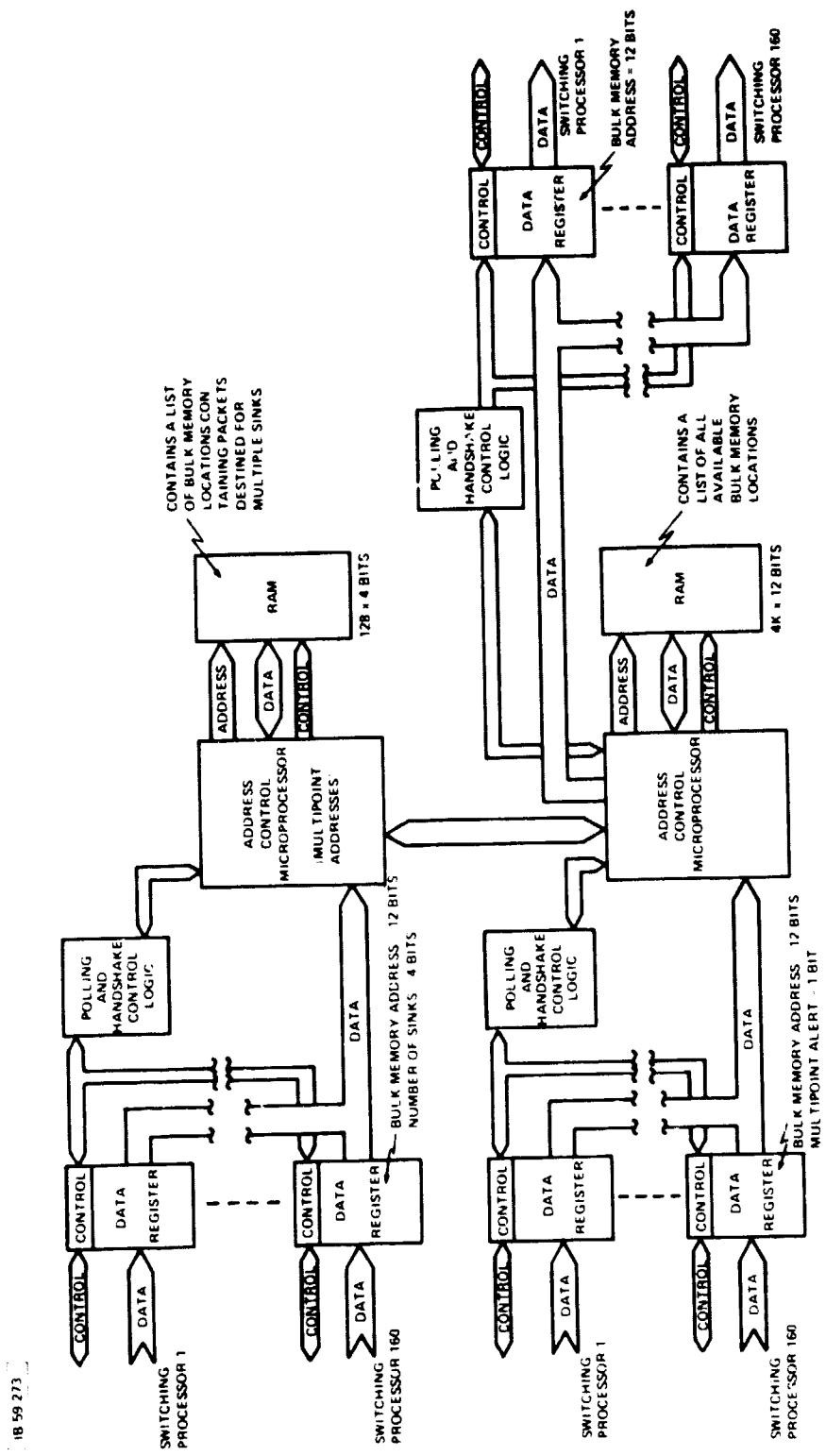


Figure 3-17. The Bulk Memory Controller

2. The last location accessed held a multipoint packet.
3. They transmitted a point-to-point packet (freeing a point-to-point BM location) and are about to receive a multipoint packet (requiring a multipoint BM location).

In summary, the switching processors are only allowed to retain the address of a BM location when the location previously held a point-to-point packet and a point-to-point packet will be stored there within the next slot.

When a switching processor returns a BM address to the BMC, it sends it to a data register in the BMC. The processor then sets a polling flag to indicate that the register contains valid data. A hardware polling circuit scans the flag of every data register in search of full registers. When the poller finds an active flag it halts, enables the output of the register and notifies the BMC control processor. The control processor fetches the address and stores it in a list containing the addresses of all available BM locations freed by the switching processors. This list resides in dedicated RAM. After the address has been stored, the control processor restarts the poller.

If the address returned to the BMC belongs to a multipoint BM location, the control processor does not store it with the other addresses. Instead, the control processor sends it to a second control processor in the BMC. The second control processor is responsible for the handling of multipoint BM addresses. This processor counts the number of times a multipoint BM address is returned. Each time the count is incremented, the processor compares the value to the number of sinks that are to receive a copy of the packet stored at that BM location.

This control processor is supplied with the number of sinks for a given packet by the switching processor that first received the multipoint packet. The switching processor sends the BM address of the packet, along with the number of sinks that are waiting for the packet, to a special data register in the BMC. The switching processor then sets a polling flag to indicate that the register contains valid BM reuse information. When the poller finds the active flag it halts and notifies the second control processor. This control processor fetches the contents from the enabled register. The number of sinks is then stored at a location in RAM that is addressable by decoding the multipoint BM address of the packet.

Each time a multipoint BM address is returned to the BMC, the second control processor accesses the value stored at the location

addresssed by the BM address and decrements it. When the value at a given location reaches zero, the BM address used to address that value is returned to the first control processor. Once returned to the first control processor, the BM address is placed into the list of available BM locations. It is now ready to be re-issued to a switching processor on demand.

When a switching processor needs the address of an available memory module, it checks its output data register in the BMC. If the register contains a valid BM address, the processor fetches this address and sets a flag to indicate the register is now empty. Otherwise, the processor must wait for the register to be filled by the BMC. A hardware poller scans the output data registers for the BMC. Whenever an empty register is encountered, the first control processor is notified. This control processor refills the register with either the address of an available point-to-point or multipoint BM location, according to what was requested. The processor then restarts the poller.

3.8.3.2 The Bulk Memory Array

The function of the bulk memory array is to hold all received packets until they can be routed and sent to the proper sink(s).

- The Input Side

The block diagram of the input side of the bulk memory array is shown in figure 3-18a. There are 160 dedicated input buffers servicing the switching processors. These buffers are attached to the input address/data bus which services the 1320 memory modules.

The total number of memory modules is based on the following assumptions:

1. The number of buffers to handle steady state traffic is:

$$BM_{ss} = (2 + \frac{28}{8}) \times 128 = 704 \text{ buffers.}$$

2. The number of additional buffers required to handle statistical fluctuations (50%) is:

$$BM_{sf} = BM_{ss} \times 0.50 = 352 \text{ buffers.}$$

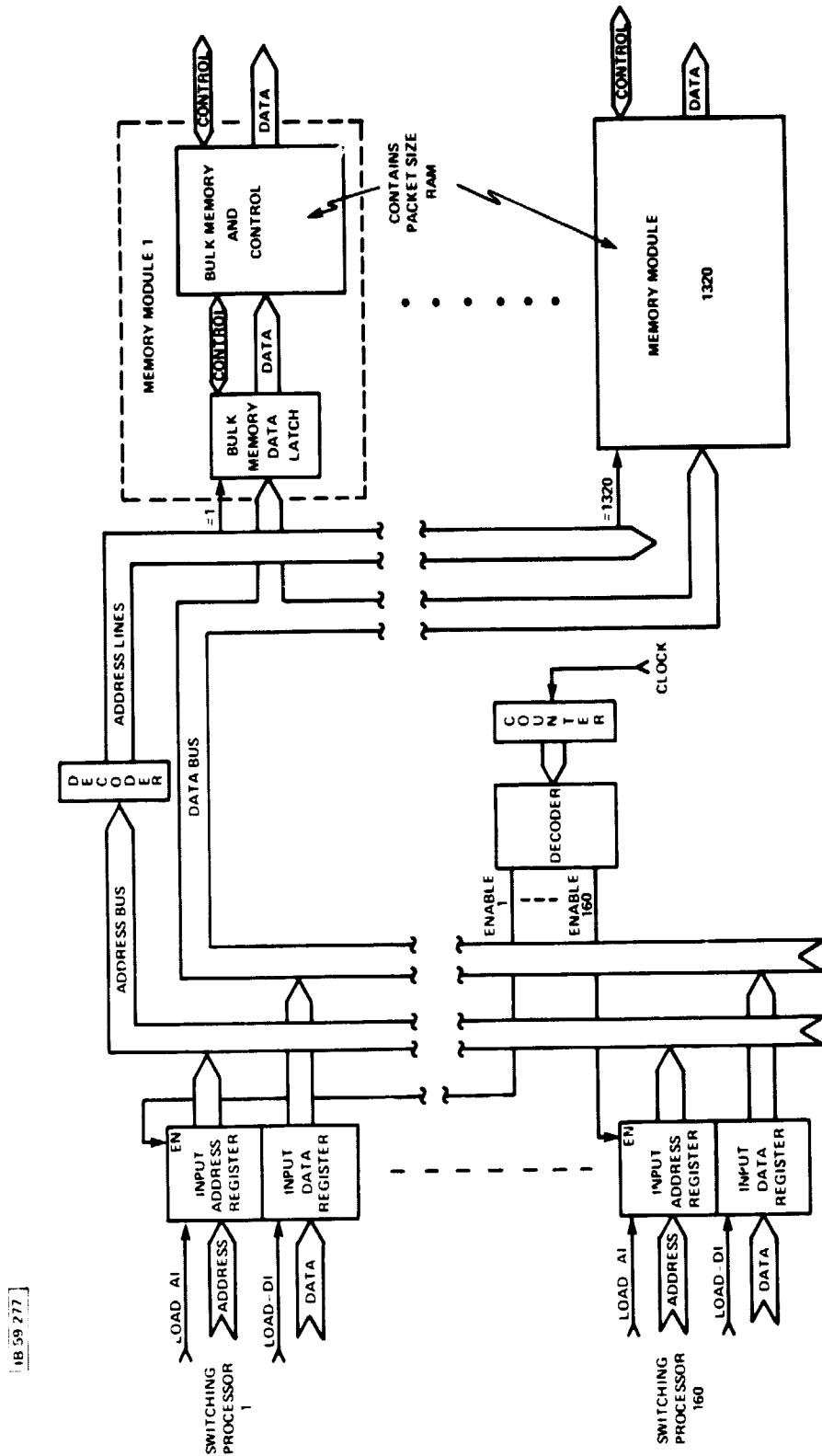


Figure 3-18a. The Bulk Memory Array (Input Side)

3. The number of redundant buffers required to achieve a 25% redundancy is:

$$BM_r = (BM_{ss} + BM_{sf}) \times 0.25 = 264 \text{ buffers.}$$

Therefore, the total number of memory modules is:

$$BM_t = BM_{ss} + BM_{sf} + BM_r = 1320 \text{ buffers.}$$

In assumption (1) the factor of 2 is due to the fact that, ideally a processor should only need a two BM locations: one to hold an outgoing packet and one to receive an incoming packet. However, since there are eight slots/frame, a packet may be forced to remain on-board the satellite for as long as eight slots (one complete frame). This leads to the factor of (28/8) found in assumption (1). The number 128 is based on the number of channels.

Assumption (2) is based on providing a 50% redundancy for safely handling most statistical fluctuations.

Assumption (3) is based on the idea that the satellite should be provided with a 25% redundancy of key hardware to by-pass hardware failures that shorten the useful life of satellites.

Of the 1320 memory modules, 132 are assigned to handle multipoint packets. Again the number 132 is based on assumptions (1) through (3) for the condition that up to 10% of the 128 channels may use the multipoint function. Thus, the number 132 is calculated as follows:

$$BM_{mp} = 10\% \times 1320 = 132 \text{ multipoint buffers}$$

The switching processors send the packets to the BM array in the form of 32-bit words which are arriving at the rate of one word per 2.6 microseconds. A hardware poller enables one input buffer onto the bus at one time. The buffer's contents include a BM address and a 32-bit packet word. The address is decoded and used to enable a single bulk memory data latch. This latch fetches the packet word from the bus in approximately 10 nanoseconds. It then sends the slower bulk memory device eight 4-bit blocks, one at a time. A block is sent to the memory device once every 320 nanoseconds, which is a reasonable access time for standard memory

devices. At the end of one slot, forty-nine 32-bit transfers have taken place for each incoming packet.

- The Output Side

Figure 3-18b contains the block diagram of the output side of the bulk memory array. On the output side of the BM array there are 160 dedicated output buffers servicing the switching processors. These buffers are attached to the output address/data bus. Whenever a switching processor is required to transmit a packet, it must place the packet's BM address into the buffer. A hardware poller enables one buffer onto the bus at one time, the address is decoded and used to select a single bulk memory output data latch. This latch places its contents (a 32-bit packet word) onto the bus. This word is then strobed into the data portion of the enabled output buffer. This data is then shifted out serially and is routed to the proper modulator by the switching processor. After the contents of the latch are emptied onto the bus, the latch fetches the next eight 4-bit blocks from the memory device. The latch is completely reloaded in 2.6 microseconds and ready to be accessed again as the poller completes its cycle. This operation continues until the packet-length shift registers in the modulators are filled and ready for the transmission cycle. This requires 125 microseconds.

If a multipoint BM module is being addressed, a slightly different operation is required during a memory read cycle. All multipoint BM locations have double output data latches. The data is held stable in one latch for the duration of one polling period (2.6 microseconds). This permits more than one switching processor to fetch a copy of the same packet during the same slot interval. While the latch being read is held stable, the other latch is being loaded with the next word. When the poller finishes one complete scan of the switching processors' output buffers, the newly filled latch is multiplexed onto the bus. The "emptied" latch is now refilled. This operation continues until a copy of the packet has been sent to every requesting switching processor.

3.8.3.3 The FIFO Controller

The FIFO controller is responsible for the routing of all packets that have been received by the baseband processor. In order to carry out this task, the FFC must be provided with the packets' destinations and their locations in the BM array. A block diagram of the FFC is shown in figure 3-19a.

After a switching processor decodes a packet's header, it sends the packet's destination address to the FFC buffer. Along with this address, the switching processor sends the BM address of the packet.

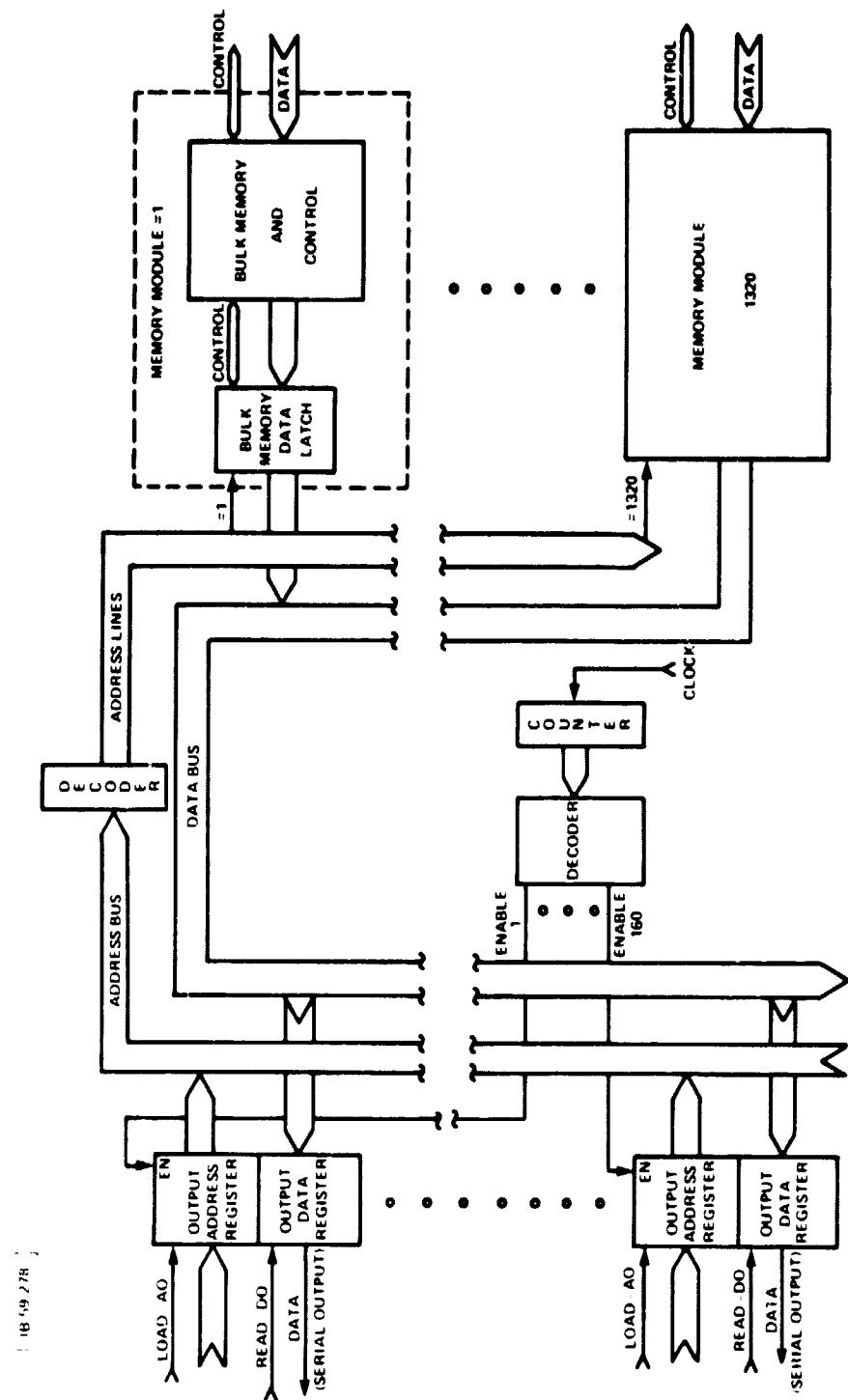


Figure 3-18b. The Bulk Memory Array (Output Side)

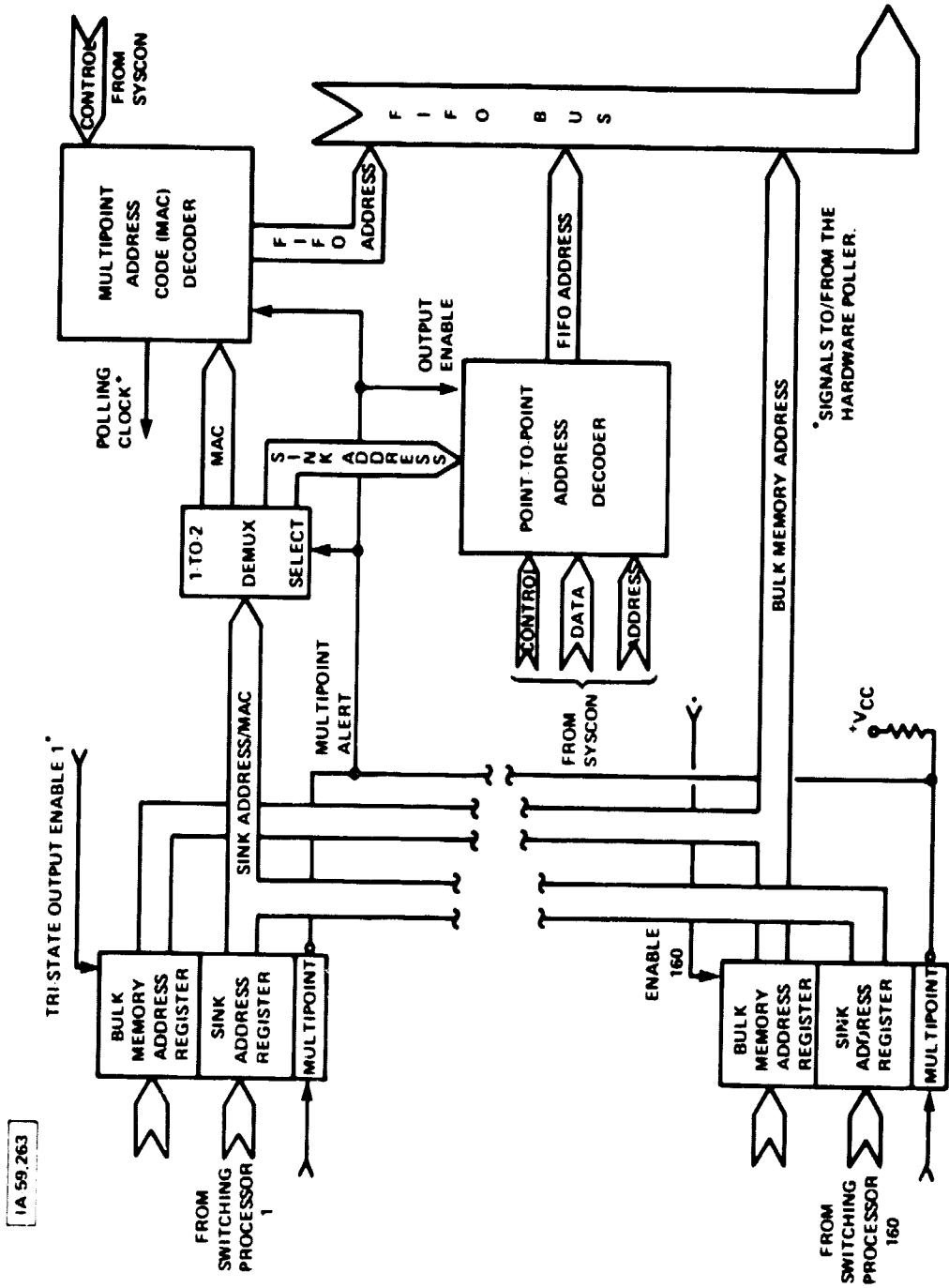


Figure 3-19a. The FIFO Controller

The processor then sets or clears the multipoint alert flag depending on the type of packet being routed. If a switching processor did not receive a packet during a slot interval, it simply clears the buffer, indicating the absence of new routing information.

A hardware poller enables the output of one of these buffers at a time. The packet's BM address is sent directly to the FIFO bus. Attached to this bus is every FIFO list in the baseband processor. If the multipoint alert flag for the enabled buffer is not set, the packet's destination address is routed to the point-to-point address decoder. The point-to-point address decoder is shown in figure 3-19b.

- Point-to-Point Routing

The point-to-point address decoder contains two address translation tables. The tables convert the actual sink addresses into FIFO list addresses. Eight unique FIFO lists are serviced by each active switching processor. In turn, each switching processor services a unique channel that broadcasts to the sinks awaiting the packets listed in the processor's FIFOs.

Due to changing traffic conditions, hardware outages, and other network-related problems, these translation tables may require periodic updating. The use of duplicate tables allows a SYSCON update of one table, while the other remains operational. This allows continuous operation without an interruption in service to the users. Also, the use of two tables provides a 50% redundancy in hardware to reduce the possibility of a fatal hardware failure. (Updates to the single surviving table would cause temporary delays in service.)

The operation of the decoder is straightforward. The sink address is used to address a location in the active conversion table. The contents of the addressed location are sent to the FIFO bus along with the already stabilized packet BM address. All the FIFO lists monitor the FIFO bus. When a list recognizes its address on the bus, it fetches the BM address. The BM address is then placed into the list in a first-in-first-out fashion resulting in the transmission of the oldest packets first. Once a packet's address has been placed into a FIFO list the packet is considered routed.

- Multipoint Routing

When the multipoint alert flag of an enabled buffer is set, the packet's MAC is routed to the MAC decoder. This decoder is shown in figure 3-19c.

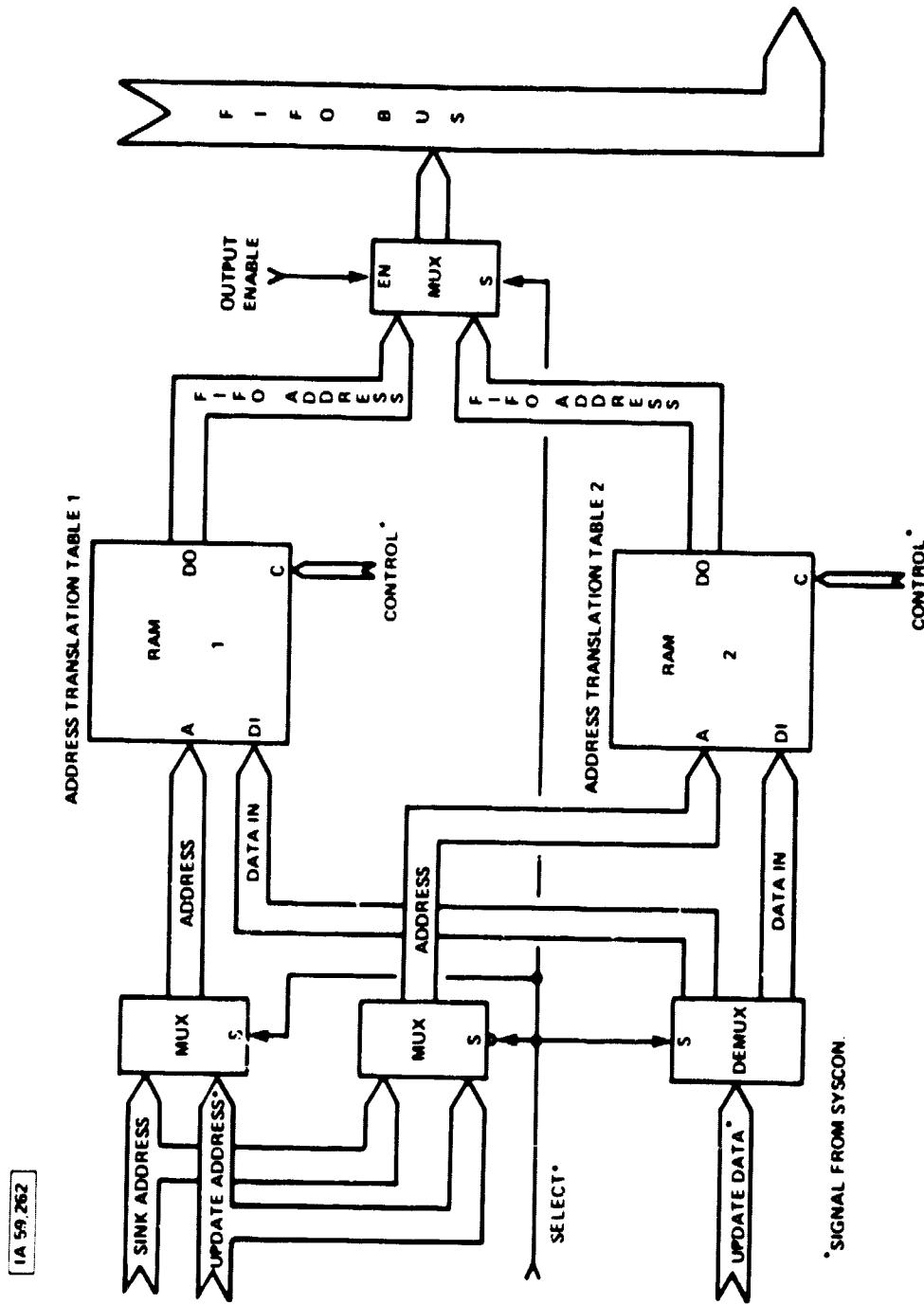


Figure 3-19b. FIFO Controller (Point-to-Point Address Decoder)

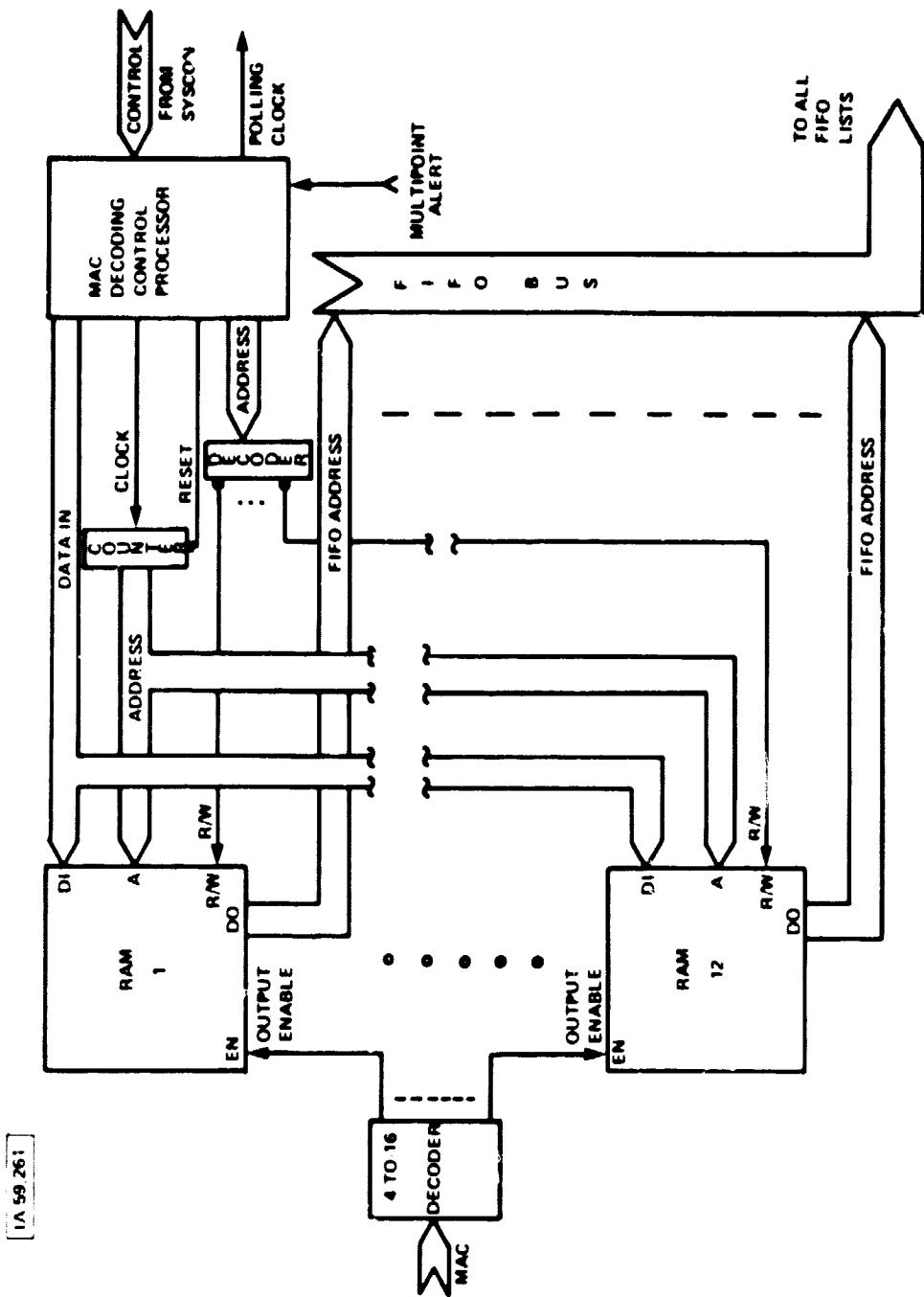


Figure 3-19c. FIFO Controller (Multipoint Address Code (MAC) Decoder)

The MAC decoder contains 12 conversion tables which each hold 16 FIFO addresses. Included in the decoder system is a MAC decoding control processor. This processor is responsible for polling synchronization for the FFC, updating the MAC conversion tables, and providing control during the MAC conversions.

During a multipoint link establishment procedure, SYSCON sends the MAC decoding processor a copy of the MAC issued to the users and the list of FIFO addresses that correspond to the current users of the issued MAC. The control processor places the list of FIFO addresses into the MAC conversion table selected for the issued MAC. Since the LEP protocol allows up to 16 sinks per multipoint call, each table holds 16 FIFO addresses. Unused locations are cleared by the processor.

Once the MAC table is prepared and the multipoint packets begin to arrive, the operation of the MAC decoder begins. The MAC is used to address the proper table. After the RAM containing the selected table is enabled, the MAC decoding processor begins to cycle through the 16 table locations. Each FIFO address leaving the RAM is sent to the FIFO bus. Here, it is held stable long enough for the addressed FIFO to recognize its address and to latch the BM address on the bus. This is done until all 16 FIFO addresses have been placed on the bus. (FIFO list address zero is not used so that table locations cleared by the control processor do not disrupt FIFO updating when placed on the bus.)

- The FFC Poller

The FFC polling circuit is presented in figure 3-19d. This circuit is clocked by the MAC decoding control processor. The function of the polling circuit is to enable one FIFO buffer belonging to an active switching processor for the duration of the decoding/routing cycle. The poller is disabled for 1 microsecond of every slot interval. During this time, every active switching processor is accessing one of its eight FIFO lists. With the FFC in a halt cycle, the switching processors can read their FIFO lists without their being updated simultaneously, which would result in unstable data.

The use of the halt cycle leaves the poller 124 microseconds in which to cycle through the 128 active FIFO buffers. However, the poller may actually have less time than this. Up to 12 of the 128 FIFO buffers may contain multipoint routing information. Each multipoint packet may have up to 16 destinations. Therefore, as many as 308 decode/route cycles may be carried out by the FFC. Each cycle must be accomplished in 400 nanoseconds.

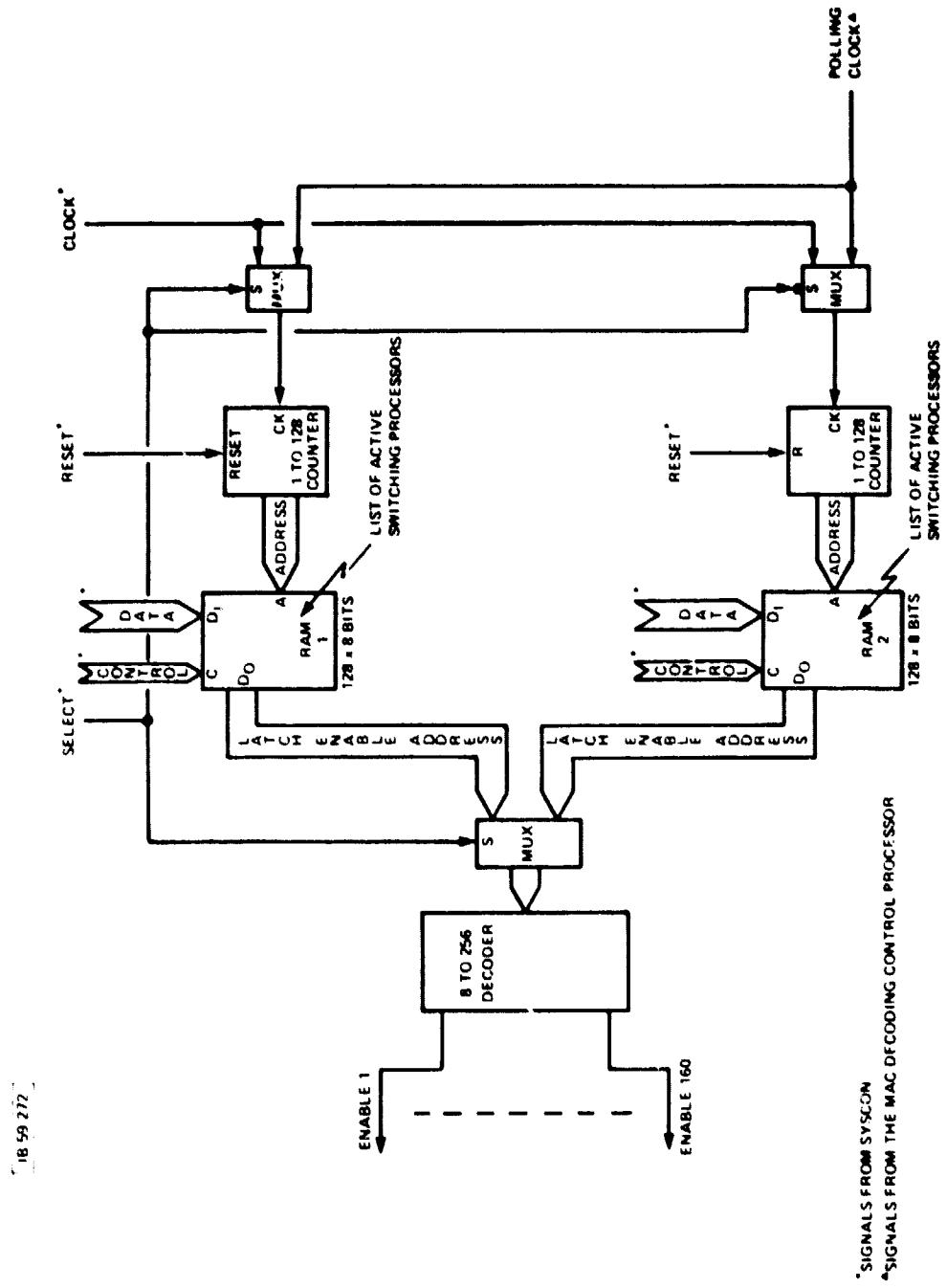


Figure 3-19d. FIFO Controller (Hardware Poller)

The MAC decoding control processor clocks the poller once every 400 nanoseconds for each point-to-point packet encountered. If the buffer just enabled contains multipoint routing data, the poller is held at that address until the 16 MAC decodings have been carried out. Since the number of decode/route cycles may range from 128 to 308 in each slot interval, the MAC decoding control processor must monitor the polling cycle. When the poller reaches the last buffer, the control processor must decide whether or not to allow the poller to start the next polling cycle. If all of the allocated 124 microseconds have been used by 308 decode/route cycles, the poller is restarted (after the 1 microsecond halt cycle for the switching processor read operation). If less than 308 decode/route cycles were required during the slot interval, the control processor holds the poller stable until it is time to start the next slot.

The FFC poller, unlike the fast BM pollers, operates at a fairly low duty cycle (even when cycling at its maximum rate). This allows the use of memory which holds a list of the active devices that require polling. Unlike the BM pollers which must poll all 160 switching processors, the FFC poller is selective. Two RAMs contain the list of active switching processors. One RAM contains the list currently in use while the other serves as backup. (The use of two lists also allows SYSCON to update the lists without interruption, whenever a switching processor assignment is changed.) Each time the polling circuit is clocked, the address of the next device to be polled is strobed out of memory. The RAM locations are addressed sequentially by a counter. The output of the RAM drives a decoder which selects the single device being polled.

3.8.3.4 The Switching Processors

A single switching processor is shown in figure 3-20. The switching processors are implemented by general purpose microprocessors, RAM, ROM, and control logic. Each processor supports its own input/output (I/O) bus. This bus permits the processors to interface with the FFC, the BMC, the BM array, and their own FIFO lists. The processors receive signals from SYSCON via a dedicated channel. Also, each processor has dedicated lines linking it to the demodulator and modulator it is serving. The processors carry out their switching tasks at a moderated rate executing a small number of instructions.

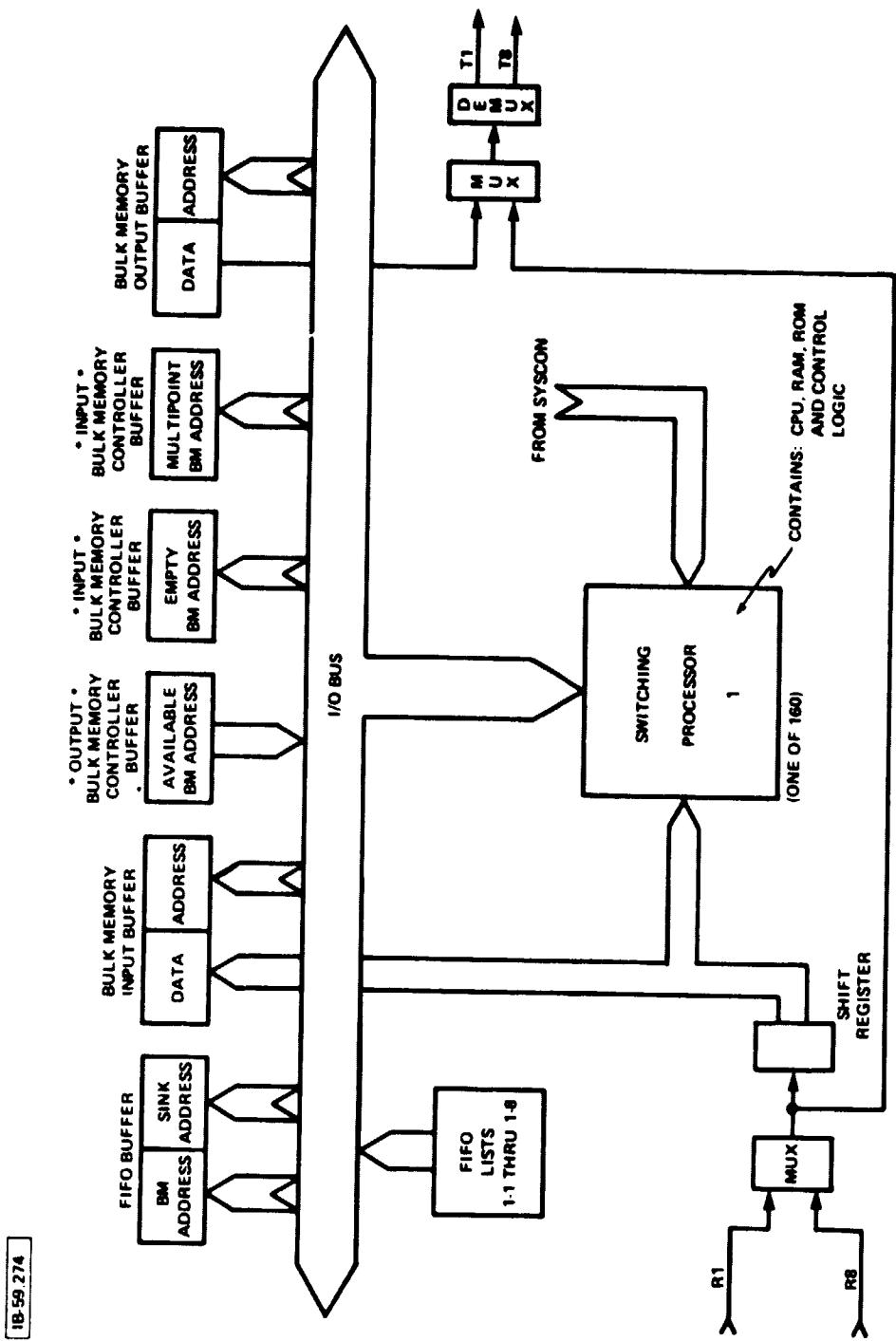


Figure 3-20. A Single Switching Processor

3.9 TECHNOLOGY IMPLICATIONS

This section summarizes a survey of both current and forthcoming electronic device technologies suitable for the design and ultimate implementation of the digital baseband processor. An update of the current state of the art for several of the more promising device technologies, with representative commercially available examples, is provided. Some of the more conservative projections of the state of the art into the coming decade, as based upon the technology development trends over the past decade, are presented. Promising new options that are now or will soon be available for inclusion into processor design are considered.

3.9.1 Baseband Processor Data Rates

Quantitative requirements to be met by the electronic devices employed in the processor design are established in a short review of the pertinent baseband data rates and times.

The proposed satellite design has 16 uplink and 16 downlink antenna beams. In each set of 16 beams, eight beams are assigned to fixed, high-traffic areas, and the other eight beams are used in a cyclical scanning mode. Each of the eight scanning beams visits eight coverage areas, providing complete CONUS coverage with a total of 64 coverage areas. With a coverage area dwell time of 125 μ s, each scanning beam has a frame time of 1 ms. Of the total channel capacity of each coverage area, a fraction equivalent to eight T1-carriers (T1: 1.544 Mb/s) will be handled by the baseband processor. Accordingly, each antenna beam will have a baseband capacity of sixty-four T1-carriers, or about 98.8 Mb/s per beam, leading to a total satellite baseband capacity of almost 1.6 Gb/s.

For the uplink, an FDMA approach was selected with eight FDM channels (of about 12.4 Mb/s per channel) for the scanned beams and eight or more FDM channels for the fixed beams. The 12.4 Mb/s channel width accommodates one T1-carrier when bursted at eight times the normal bit rate since the coverage area dwell time is one-eighth of the frame time.

For the downlink, a TDM broadcast approach was selected to carry the 98.8 Mb/s beam capacity which results from bursting eight T1-carriers (per coverage-area dwell) at eight times the normal T1 rate. That is, each of the eight output channels routed by the baseband processor to a given downlink beam modulator has a data rate of 12.4 Mb/s. If the output channels are byte-wide (8 bits) data busses, then the output channel data rate is 1.544×10^6

(1-byte) words/s. If two 64-bit shift registers, operated in ping-pong fashion, are alternately parallel loaded with eight 8-bit words every $0.648 \mu\text{s}$ ($= 1/f_{T1}$), they can alternately be serially shifted out to the modulator at the desired 98.8 Mb/s rate.

3.9.2 Major Areas Requiring State-of-the-Art LSI Technology

There are three major areas in the design of the satellite-borne baseband digital processor which depend heavily on state-of-the-art LSI technology:

- Demodulation of the FDMA uplink beams into serial bit or byte data streams
- Baseband digital processing of these data streams
- Remodulation of the processed and routed data streams into TDM downlink beams

Consideration of some of the major aspects of these areas serves to introduce the specific devices which are dealt with in later sections. Since there are 16 beams, each with at least eight frequency division multiplex (FDM) channels, there will be at least 128 input (and 128 output) channels in the baseband processor. One should thus be mindful of this high degree of channel replication multiplicity when choosing a specific device to implement a channel function.

3.9.2.1 LSI Requirements in Demodulation and Remodulation

After demodulation of the baseband component of each uplink beam, each of the eight-channel demodulators will deliver a 12.4 Mb/s serial data stream. By using an 8-bit wide serial-to-parallel conversion shift register, the channel data bus delivers one byte every $0.648 \mu\text{s}$. Prior to remodulation, a 64-bit wide parallel-to-serial conversion shift register is used to produce the 98.8 Mb/s data stream from the one byte per $0.648 \mu\text{s}$ outputs of the eight channels routed to the given remodulator.

As to the demodulation and remodulation processes themselves, several attractive analog and digital methods can be considered. Among the analog approaches are use of charge-coupled devices (CCDs), surface acoustic wave devices (SAWDs), and transferred electron logic devices (TELDs). Digital demodulation involves analog-to-digital conversion and (sometimes) sample-and-hold circuitry. Digital remodulation involves digital-to-analog conversion and perhaps some filtering. Both digital demodulation

and digital remodulation require byte-by-byte multiplication of data with local oscillator signals, whether explicitly by combinatorial-logic parallel multipliers and associated accumulators, or implicitly via, e.g., the CORDIC rotation algorithm implemented on an IC chip.

3.9.2.2 LSI Requirements in Baseband Processing

Depending upon the level of functional complexity desired, the baseband processor design will include some (or all) of the following representative features:

- Header Extraction/Processing
- Traffic Routing (according to header information)
 - Data channel multiplexing/demultiplexing
 - Queueing
 - Traffic prioritization
- Interim (i.e., intra RCV/XMT) Data Storage
 - Down-link scanning beam delay
 - High-channel-rate queueing
 - Low priority traffic
- Error Management
 - Header bits - FEC
 - Packet data - ARQ (with processor-recalculated CRC word)
- Encryption of (Broadcast) Downlink Transmission

3.9.2.3 Power -- Dominant Criterion?

Although the size of the channel memory and the high-level performance of the processor logic, together with the speed of both, are clearly very important considerations, the limited power available on board the satellite and the rather large number (≥ 128) of channel replications cause device power to assume almost dominant importance.

If, for example, 1.28 kW (so as to yield a round-number final value) is available for demodulation, baseband processing and remodulation, then a power-budget of 10 W/channel results. It is this number against which a candidate device or component must be compared to assess feasibility and appropriateness of incorporation into channel design. Thus, unless all other components are exceptionally low power devices, a fast multiplier approach consuming 5W cannot be reasonably justified.

3.9.3 Overview of LSI Technology

Three categories of LSI technology development are of direct interest to the design of the satellite digital baseband processor. The first concerns technologies which are currently well advanced and which may, where suitable, be relied upon here and now. It is these technologies that form the basis for the technology projections into the future. The second category concerns technologies with near-term promise (mid 1980's?). Reasonable and conservative extrapolation of the results of the past decade suggests that the expected products of these technologies are worth tentative incorporation into the baseband processor design. The third category concerns technologies with far-term promise (early 1990's?). These technologies, extremely attractive in what they promise, are still in their infancy and extrapolation of current laboratory results into the next decade is divination at best. Although such technologies cannot be relied upon for current designs, neither can they be ignored.

3.9.3.1 Presently Well-Advanced Device Technologies

Devices of interest are mainly LSI-level logic and memory chips although a few special processing chips are also considered. At present, silicon devices dominate the integrated circuit (IC) picture, but inroads are being made by gallium arsenide in the area of low-power gigabit/second devices. Also, thin magnetic bubble domains are entering into the very-high-density memory picture as are compounds such as quartz and lithium niobate in signal processing.

Integrated circuits fabricated with silicon utilize both bipolar and field-effect transistor technologies, but gallium arsenide integrated circuits use only the field-effect transistor technology.

Bipolar Devices

Bipolar devices, whose transistor action is associated with pn-junctions and minority-carrier current dynamics, are generally the faster of the available devices, but also are the highest in power dissipation. Having the greatest radiation hardness, bipolar ICs form the main military technology. Some of the more common bipolar transistor logic families are described below.

TTL (Transistor-Transistor Logic). Probably the most prevalent bipolar logic family today, TTL technology is used in a wide variety of small scale integration (SSI)- and medium scale integration (MSI)-chip logic functions where its relatively high power dissipation is tolerable. In addition, for these applications as well as for memories in particular, lower-power (L) and faster, Schottky-diode-clamped (preventing collector saturation) (S) versions of the TTL technology more suited to LSI are often available. A typical, currently available TTL 4k RAM has a 30 ns access time and dissipates about 850 mW (about 210 μ W/bit). A megabit of this type of memory would accordingly dissipate over 200 W.

I²L (Integrated Injection Logic). This form of bipolar logic lends itself well to LSI with its high packing density (150 gates/mm²) and low power/delay-time product (~1 pJ). I²L is a low voltage (1-V supply, 0.6-V transitions) current-steering technology. The base currents of several gate input transistors, supplied by the collector of a single transistor acting as a constant-current source (injection-current bus), can be shorted out (or left alone) by the gate input signals. For low to moderate injection current levels, the switching response time is inversely proportional to the power dissipation (controlled by the injection current level), which provides a useful speed-power trade-off design option. IBM has recently reported significant improvement in their I²L technology. Using 2.5 μ m design rules, they have achieved average switching speeds of 0.8 ns (or a maximum toggle frequency of 625 MHz) at 100 μ A per gate which translates into a dissipation of approximately 100 μ W per gate, yielding a speed-power product of about 0.08 pJ.

ECL (Emitter Coupled Logic). ECL is the fastest of all the silicon logic families and also dissipates the most power. The heart of the ECL gate is an emitter-coupled differential amplifier in which several parallel, common-collector input transistors drive the output common-base transistor. The collector loads are selected so that no collectors enter saturation. This provides very fast (typically sub-nanosecond) switching action but at the price of relatively high power dissipation per gate. Typically, very fast bit-slice architecture designs employ ECL. A typical, currently

available ECL 1k RAM has a 12 ns address access time and dissipates about 500 mW (500 μ W/bit). A megabit of this type of memory would accordingly dissipate 500 W. Thus, ECL must be used very selectively, if at all, in this satellite application.

FET Devices

FET devices are based on field effect transistor action in which potentials on the high-impedance gate electrode modulate the flow of the majority-carrier current between the source and drain electrodes. The method of gate isolation employed gives rise to two major categories of devices. In one category, devices whose gates are isolated from the conduction channel by means of a reverse-biased pn junction are known as junction field-effect transistors: JFETs or simply FETs. JFETs are called MESFETs when the reverse-biased isolation junction is a metal-semiconductor rectifying contact (i.e., Schottky barrier).

Devices in the other category, whose thin-film metallic gate electrodes are isolated from the (semiconductor) conduction channel by means of a thin layer of dielectric insulation, are in general called either MISFETs or IGFETs: metal-insulator-semiconductor or insulated gate FETs. When, as is most commonly the case, the dielectric material is an oxide, the devices are then called MOSFETs: metal-oxide-semiconductor FETs.

Among FET devices, there is an important further dichotomy based on the nature of the conduction channel. The FET whose channel is fully conductive at zero bias (e.g., a channel diffused or ion-implanted into the substrate), and is made less conductive (depleted) by reverse biasing the gate, is called a depletion-mode FET or sometimes DFET. FETs, whose channel is unformed at zero bias, and is created (with forward gate biasing) by inducing an inversion layer near the oxide-semiconductor interface, are called enhancement-mode FETs or sometimes ENFETs.

FET devices dissipate substantially less power than bipolar devices but are generally somewhat slower, although the speed gap is narrowing (and sometimes even closed). Some of the more common FET logic families are described below.

PMOS (p-Channel MOSFETs). PMOS technology was first to be developed because most of the contaminants in MOS fabrication are mobile ions which are positively charged and are easily trapped in the oxide layer that insulates the gate from the substrate. In p-channel devices, these positive contaminant ions, due to negative gate biasing, are collected at the metal (gate)-oxide interface and have little effect on the channel. In n-channel devices, however,

the positive contaminants collect along the oxide-semiconductor interface and induce a shift in bias which causes the transistor to turn on prematurely. The more extensive process control required to fabricate comparable n-channel devices initially made them uneconomical. The intrinsically superior performance of n-channel devices, however, prompted research in fabrication technology which ultimately resulted in PMOS logic being supplanted by NMOS in all but the simplest and least-expensive consumer products.

NMOS (n-Channel MOSFETs). NMOS logic is superior to PMOS because the electron mobility in silicon is more than 2.5 times the hole mobility; this ultimately translates into a speed advantage for NMOS. For the same operating conditions, PMOS resistivity will be 2.5 times higher, requiring 2.5 times higher areas to achieve the same resistance. Accordingly, NMOS packing densities can be made larger, for the same complexity, and thus faster in switching due to the smaller junction areas and resultant smaller capacitances that figure into the internal RC time constants that directly limit device operating speed. A representative state-of-the-art NMOS microprocessor, the MOTOROLA MC68000, has some 13,000 gates (with over 70,000 active elements) on an approximately 6mm x 7mm chip using 3.2 μ m minimum features, and dissipates 1.2 W at an 8 MHz clock rate. A representative NMOS static RAM of 16 kbit capacity with 80 ns address access time dissipates about 190 mW (11.9 μ W/bit.) One megabit of this type memory would then dissipate about 12 W.

CMOS (Complementary MOS). In complementary MOS logic, both NMOS and PMOS enhancement mode transistors are used on the same chip to produce a logic family whose standby (i.e., inter-logic-transition) power dissipation is extremely low. The use of both channel polarities is achieved by diffusing or ion-implanting wells of polarity opposite to that of the substrate, effectively creating local islands of opposite polarity substrate. Thus, with an n-type substrate, PMOS transistors are made on the direct substrate, and NMOS ones are made on the p-type islands or wells.

If PMOS and NMOS transistors are stacked in series between a power supply and ground with gates tied together, there is no appreciable quiescent current (nor, therefore, power dissipation) through the stacked transistors since only one or the other can conduct. If there is stray capacitive loading at the inter-drain tie-point between the two transistors, one or the other transistor will supply a brief charging current pulse to alter the voltage across the stray capacitance. Thus CMOS logic is characterized by extremely low steady-state power dissipation and dynamic power dissipation that increases with increasing clock rate. Another important benefit of complementary MOS logic is that of equal drive

power for either polarity transition versus, for example, active pull-down and passive pull-up common in other logic families. In addition to its low power dissipation qualities, CMOS possesses the further advantages of high noise immunity, wide tolerance to power supply variation and low temperature sensitivity.

Despite its lower static power dissipation (previously not too important) CMOS has trailed NMOS in density and speed development because the CMOS process is somewhat more involved. CMOS is usually considered to be slower than NMOS when its less well-developed density, and therefore speed, is not taken into consideration. Actually, however, for a given patterned gate line width, CMOS is about twice as fast as NMOS. With the drive to very large scale integration (VLSI) and greatly increased interest in minimizing power dissipation, CMOS development is advancing rapidly.

A representative CMOS microprocessor, the RCA CDP1802, has an 8-bit internal data bus, a sixteen 16-bit-register general purpose register file, and a 91 instruction repertoire. With a 5 V supply, it dissipates 0.5 mW quiescently and 6 mW at a 3.2 MHz clocking rate; or, with a 10 V supply, it dissipates 5 mW quiescently and 40 mW at a 6.4 MHz clocking rate. This contrasts dramatically with a typical (though somewhat more powerful) NMOS microprocessor such as the INTEL 8085A, which dissipates 850 mW at a 5 MHz clocking rate. Also, a representative CMOS 16-kbit static RAM with 100 ns address access time dissipates 200 mW (12.5 μ W/bit) at full cycle rate, but only 1 mW (0.063 μ W/bit) quiescently. Thus, a 1 Mbit CMOS memory would dissipate 12.5 W at full cycle rate but only 63 mW quiescently.

CMOS/SOS (Silicon on Sapphire). The previously described type of CMOS logic is known as bulk CMOS since it is fabricated on a silicon substrate, into which the various transistor elements are diffused or ion implanted. These elements are DC-isolated from the bulk substrate since they are biased so as to form reverse-biased pn-junctions. They are, however, capacitively coupled to the (semiconductive) substrate through the pn-junction depletion region capacitance. This stray capacitance serves to reduce the switching speed. If a thin film of n-type silicon is hetero-epitaxially deposited on a dielectric substrate of sapphire, however, with transistor elements formed by polarity-reversing amounts of diffused or ion-implanted doping, then the capacitive coupling of transistor elements to the substrate is greatly reduced. The reduced capacitive loads permit lower current-drive designs and thereby smaller and more densely packed transistors.

This capacitive advantage is somewhat offset by the reduction of the majority carrier mobility from the bulk value to the surface

value, which results, in turn, in a reduced transistor gain constant. Also, the standby leakage current in currently fabricated CMOS/SOS is much higher. Although CMOS/SOS is at present a more difficult and expensive technology, it holds a substantial advantage in gate power, and is expected to play a significant role in the advance to VLSI. Currently available CMOS/SOS static RAMs include an RCA 4 kbit one with 650 ns access and 0.25 mW (0.063 μ W/bit) power dissipation, and a 16 kbit one (recently announced) with 80 ns access and 14 mW (0.88 μ W/bit) dissipation at full cycle rate and 5 μ W quiescent dissipation. With the latter type CMOS/SOS memory, a 1 Mbit memory would dissipate 880 mW at full cycle rate and 310 μ W quiescently.

MNOS (Metal-Nitride-Oxide-Semiconductor). When the previously treated logic family technologies are used to fabricate memory arrays, the memories thus formed are all volatile: they lose all information content upon powering down. Nonvolatility can be introduced into a semiconductor memory through the mechanism of charge trapping. One of the earliest approaches to realize this replaced the simple oxide gate-insulation layer with a composite oxide-nitride layer. A very thin silicon dioxide layer (e.g., 20 Å) was deposited on the substrate over the FET channel area, followed by a thicker silicon nitride layer (e.g., 600 Å), and in turn by the conductive gate-electrode layer. This nitride layer contained charge-trapping centers which would immobilize any charges that could be brought to them.

By application of relatively high potentials (25-30 V) to the gate electrode, charges from the substrate could be induced, by quantum mechanical (Fowler-Nordheim) tunneling, to traverse the thin oxide layer and become trapped in the nitride trapping centers near the oxide-nitride interface. High reverse-polarity potentials could reverse this process. The trapped charges exert the same effect on the FET channel below them as a potential permanently applied to the gate electrode. For instance, if the trapped-charge potential strongly off-biases the FET, then application of a normal-level gate strobe would fail to make the FET conduct, and vice-versa in the absence of any trapped-charge off-bias.

Use of the trapped charge mechanism to achieve nonvolatility introduces two further memory parameters: retention and endurance. Retention relates to how long (typically years) the trapped charge takes to leak away when there have been fewer than, e.g., 10-100 erase/rewrite cycles. Endurance relates to how many (typically 10-100 k) erase/rewrite cycles it takes to "stress-fatigue" the nitride layer so that the charge/no-charge thresholds are no longer different enough to provide error-free memory readout.

Due to the relatively limited number of erase/write operations inherent in MNOS-based memories, they are generally referred to as EAROMs: electrically alterable read-only memories. Currently available EAROMs are relatively slow in readout times (typically 1 μ s or more) -- and extremely slow in erase/write cycle times (typically 10-100 ms). Most presently require several power supplies, although this may be remedied in newer designs. They do permit both single-word as well as block erasure, and, what is most important, allow it to be done within the circuit. A representative EAROM has a 4 kbit capacity, 900 ns read access time, 10 ms erase/write cycle, 10 yr retention, 10^5 cycle endurance, and requires +5 V and -12 V supplies, dissipating 450 mW when operating.

FAMOS/FLOTOX. In addition to the MNOS approach to nonvolatile memory, there are processes called FAMOS: floating-gate avalanche-injection MOS, and FLOTOX: floating-gate tunnel oxide. As their process names indicate, the gate electrode is electrically floating, surrounded by the silicon dioxide gate insulation material. There are several mechanisms by which charge can be made to traverse the oxide layer and become stored on the floating gate electrode. In one approach, back-biasing the drain-substrate junction with a high (e.g., 25-30 V) voltage initiates an avalanche discharge and some of the resultant high-energy electrons burrow through the thin oxide layer to the floating gate electrode. In another approach, producing an intense electric field between a control gate electrode (above, and oxide-insulated from, the floating gate) and the substrate enables electrons to traverse the oxide layer by quantum mechanical tunneling. In either event, the charges stored on the floating gate electrode have an effect on the FET thresholds similar to the nitride-trapped charge effect in the MNOS case.

When the resulting memory device is provided with a quartz window and is erased by ultra violet (UV)-induced photoconductivity within the oxide layer, the device is called either a UV-EPROM or just EPROM: erasable programmable ROM. When erasable by a reverse-polarity electric field, the memory device is called an EEPROM: electrically erasable PROM. In the interest of completeness, it might also be pointed out that the term ROM, read-only memory, beside denoting the generic memory category also implies factory (mask) programming; and the term PROM, programmable ROM, similarly denotes the generic memory category and implies user or field (fused link) programmability as well.

A representative (industry standard) 16K UV-EPROM has a 350-450 ns access time and requires only a single +5 V supply, dissipating 550 mW when operating. By UV illumination, block (but not single byte) erasure is effected in 15-30 minutes. Although the range of EEPROM devices is not yet as wide as UV-EPROMs, their

access times and retention are comparable. Their erase time is much shorter, a minute or less, but present devices use special voltages and complex voltage sequencing, making them impractical for in-circuit programming. They are also block-only erasable.

Gallium Arsenide (MSI)

Semiconductor technology has been dominated by devices fabricated from silicon. The industry has progressed from the single transistor to the monolithic integrated circuit and then on to large scale integration of thousands of devices on a single chip. The widespread use of silicon ICs is attributable for the most part to the industry's achievement of nearly doubling the density of operational functions per chip each year, and at a lower cost per function. As a result, applications of silicon LSI chips are growing at an enormous rate, and a new revolution in the form of VLSI is rapidly becoming possible, with subsystems, and even entire systems, placed on a single chip.

Emerging from the background of the silicon LSI/VLSI revolution, however, gallium arsenide (GaAs) is gradually gaining momentum. It may well become the major contender to the dominance of silicon technology, with potential gigabit circuitry having gate switching times of less than 100 ps and dynamic switching energies of less than 0.1 pJ. Consequently, GaAs is of steadily mounting importance for its potential in the VLSI picture.

Application of silicon ICs, particularly LSI chips, has so far been limited to relatively low data rates and has yet to invade the microwave regime. Various silicon IC technologies, including CMOS/SOS, NMOS and (bipolar) ECL, generally have gate propagation delays of 2 ns or more. Except for a few small-scale ECL applications (e.g., high speed prescalers), clock rates and function execution times are well under gigabit rates. In contrast, however, GaAs digital ICs made with depletion-mode MESFETs and Schottky diodes have demonstrated gate propagation delays of less than 200 ps, breaking the gigabit barrier and moving directly into the microwave area. The low-power, high speed advantage of GaAs MESFET ICs stems directly from the high electron mobility and semi-insulating substrate of the GaAs MESFET, which gives it high transconductance and unity current gain bandwidths of nearly 80 GHz for 1 μ m gate devices (compared with about 12 GHz for a similar Si MESFET). These same properties, of course, are the basis for the current attention being given to the development and use of discrete microwave GaAs FETs.

Since the principal advantage of GaAs is its high electron (but small hole) mobility, a feature best exploited in a majority-carrier

device technology, most GaAs IC efforts have been based on the use of n-channel FETs of various types. Of these, the Schottky diode FET logic (SDFL) structure, which uses high speed Schottky diodes for most logic functions, and 1 power depletion mode MESFETs for inversion and gain, holds great promise for LSI circuit application. SDFL achieves speeds ($t_d = 75$ ps) close to those of buffered FET logic (BFL), in which all logic functions are performed with active DMESFET elements, but at much lower power levels ($P_D = 200\text{-}2000 \mu\text{W/gate}$), promoting chips of much higher complexity. SDFL does, however, require a more sophisticated fabrication approach in order to optimize both diodes and FETs.

Two independent studies, by the Rockwell International Science Center and the IBM Thomas J. Watson Research Center, projected just what the speed advantage would be if LSI circuits were fabricated from GaAs instead of silicon, using MESFETs. The research was tackled differently by each: IBM used computer simulations for its projections, while Rockwell based its projections on actual device measurements. However, both arrived at essentially the same result: all things being equal, GaAs MESFET LSI circuits exhibit a sixfold speed advantage over silicon MESFET circuits for the same power-delay product; GaAs circuits exhibit a 25 to 40 times lower power dissipation than silicon for the same t_d ; and GaAs ICs operate at up to 200°C without changes in T_{AA} or P_D .

Gate delays and power-delay products of the GaAs enhancement-mode MESFETs are orders of magnitude better than those of today's Si n-channel MOSFETs, where the GaAs devices are formed with gate widths of $1 \mu\text{m}$ and less. One company very active in GaAs LSI research, Rockwell International, expects gate delays of 60 ps and power-delay products of 35 fJ in its $1 \mu\text{m}$ MESFET technology for logic swings of one volt. The Rockwell program has the aggressive goal of putting at least 1000 gates on a GaAs substrate within the year, and integration densities will most certainly not stop there. They have fabricated and tested, among other things, a 24-gate, 3-stage binary ripple counter (i.e., divide by eight) implemented with T(toggle)-connected D flip-flops. Power dissipation per gate for MSI circuits such as this have been running in the few-milliwatts range. The first divide-by-two stage of the counter has been operated at a clock frequency of 1 to 2 GHz, implying gate switching times of approximately 100 ps. The few-milliwatts power dissipations yield $P_D t_d$ products of a few hundred fJ/gate, which is much lower than that achieved with somewhat faster BFL. In the second quarter of 1980, Rockwell's dual 32-bit shift register, containing 532 gates, should be available.

The Rockwell International Thousand Oaks Lab (CA) has reported a 75-gate 3x3 parallel multiplier exhibiting a 172 ps gate delay.

Operating at a gate dissipation of $570 \mu\text{W}$ leads to a speed power product of 128 fJ ; reducing the power to $420 \mu\text{W}/\text{gate}$, a gate delay of 225 ps results (and a 94 fJ speed-power product). These results were extrapolated to a 1000-gate 8×8 multiplier with a 6 ns multiply time -- dropping to 2 ns with further layout improvements, at a total power consumption of about 300 mW . They hope to demonstrate the 8×8 multiplier within the year (1980). Rockwell believes that 10^4 gate densities may be achieved by about 1983.

A Philips Research Lab in France has recently reported a BFL divide-by-two scaler working at 4.4 GHz , and a complementary clock generator working at 5 GHz .

In assessing the performance results achieved by GaAs ICs relative to current silicon devices, it should be remembered that the GaAs devices fabricated to date use $1 \mu\text{m}$ or less geometries. This leads some to ascribe the outstanding GaAs performance results more to the $1 \mu\text{m}$ geometries and semi-insulating substrates (similar to SOS parasitic capacitance reduction for silicon CMOS) than to any superior GaAs electronic properties and FET characteristics, regardless of the Rockwell/IBM studies.

Although the DMESFET-based SDFL structure presently enjoys wide application, it by no means forms the complete GaAs picture. There are a number of GaAs devices, all n-channel FETs, since the low hole mobility of GaAs precludes the use of p-channel or bipolar devices, and even a complementary logic approach similar to CMOS in silicon does not appear promising. The depletion-mode Schottky barrier FET (DMESFET) is the most widely used GaAs device, and also the one that has given the highest performance to-date. Circuits employing DMESFETs pose the fewest fabrication problems, but necessarily require two power supplies. The logic gates must contain some form of voltage level shifting which imposes some penalty in terms of wafer area utilization.

The enhancement-mode GaAs FET (EMESFET) offers circuit simplicity because the logic gates require only one power supply, but the permissible voltage swings are rather low because the Schottky barrier gates on GaAs cannot be forward biased above 0.6 - 0.8 V without drawing excessive currents. Although a half-volt swing is quite reasonable for the operating range of fast, ultra-low power circuits, very tight control is required in order to fabricate uniform, very thin active layers that are totally depleted at zero gate bias and yet give good device transconductance when the device is turned on.

Larger gate voltages can be handled by substituting a pn-junction for the metal-semiconductor (Schottky) junction to form a junction gate FET (JFET). The larger built-in voltage of the pn-junction allows the GaAs JFET to be biased up to $V_{gs} \sim 1$ V without excessive conduction. Fabrication, however, presents a more difficult processing problem and thus GaAs development for JFETs is at a less well-developed stage than for MESFETs. A gate voltage even larger than the JFET's can be obtained by making the gate from a p-type semiconductor with a band gap larger than GaAs, forming a heterojunction gate FET (HJFET). With the p-type $Ga_{0.5}Al_{0.5}As$ alloy, the resulting HJFET can be biased up to $V_{gs} \sim 1.4$ V without drawing significant current. Again, processing poses a difficult problem.

Implementation of a MISFET or even MOSFET technology in GaAs would eliminate the logic swing limitation completely. So far, however, attaining such devices has proven quite difficult; the Japanese are still vigorously pursuing this goal.

GaAs technology trade-offs are listed in table 3-4.

Using the GaAs devices described above, three major GaAs logic circuit families have evolved. Buffered FET logic (BFL) is the oldest, fastest and most power hungry. Direct-coupled FET logic (DCFL) dissipates the lowest power but also exhibits the slowest speed. Lastly, Schottky-diode FET logic (SDFL) steers a middle course between the previous two approaches but still achieves the lowest speed-power product.

GaAs logic family performance results (operated in cascaded-gate ring oscillator fashion) are listed in tab' 3-5.

A recent advance in GaAs materials research by the Office of Naval Research (ONR) may have identified both the cause and the cure for some performance and reliability problems in GaAs devices. The mechanical instability of impurity ions in the abrupt dopant gradients necessary for high gains at high frequencies have been traced to induced strain fields in the GaAs die which enhance the mobility of dopant impurities so that they more easily diffuse from the surface into the bulk material, thereby altering the intended dopant profile. The suggested cure is to dope the GaAs with impurities that cluster into aggregates too large to migrate through the lattice. One manufacturer has already applied this ONR research to building devices with the steep inpuirty gradients necessary for X-band frequency operation.

Table 3-4
Trade-Off Between GaAs IC Technologies

Active FET Device	Mode	Circuit Complexity	IC Fabrication	Can Be Planar?	Potential Yield
MESFET	D	2 Supplies	Simple	Yes	High
MESFET	E	1 Supply	Doping Very Critical	Yes	Low
JFET	E	1 Supply	Requires pn-Junction	Yes	Medium
HJFET	E	1 Supply	Requires epi-growth	No	Low

Table 3-5
Comparison of Ring Oscillator Performance
for the Main GaAs Logic Families

GaAs Logic Family	Width (μm)	Gate Length (μm)	Power Dissip. (mW)	Propag. Delay (ps)	$p_D t_d$ (pJ)	Fan In	Fan Out
BFL (D-mode)	20	1.0	40	86	3.9	2	2
BFL (D-mode)	10	0.5	5.6	83	0.46	1	1
	50	0.5	41	34	1.4	1	1
SDFL (D-mode)	5	1.0	0.17	156	0.027	2	1
			0.48	107	0.051		
	10	1.0	0.34	120	0.040	2	1
			0.73	95	0.069		
	20	1.0	1.10	99	0.087	2	1
			2.26	75	0.170		
DCFL (E-mode)	20	1.2	0.10	300	0.03	1	1
	20	1.2	0.10	430	0.10	2	2

Charge-Coupled Devices (CCD)

Another very interesting type of semiconductor LSI storage device, with a wide range of applications, is the charge-coupled device (CCD), which is an example of a still broader class of structures called charge-transfer devices. These are dynamic devices in which packets of charge are moved in shift register fashion along a closed path under the control of phase-staggered clock pulses. The major portion of the charge packet transfer path consists of a periodic linear array of MOS capacitors. Depending on whether a bi- or tri-phase clocking scheme is used, every second or third MOS capacitor "gate" electrode is connected to one of two or three common clocking pulse lines.

Basic to the operation of the CCD is the dynamic storage and withdrawal of charge in a series of MOS capacitors. When a large positive pulse is applied to an MOS gate on a p-type substrate, a depletion region is created beneath the gate electrode, accompanied by a considerable increase in the surface potential under the gate. Thus, the surface potential "depression" forms a potential well which can serve as a container for stored charges.

If a constant positive gate bias is applied for a sufficiently long time, electrons thermally generated near the substrate surface will accumulate at the surface, forming a steady-state inversion layer. The amount of inversion charge accumulated is a measure of the capacity of the well for storing charge, and the time to fill the well is called the thermal relaxation time, which depends on the quality of the semiconductor material and its interface with the oxide insulator. For good (negligible charge loss) CCD operation, material quality should be high enough that the thermal relaxation time is much longer than any (non-regenerated) storage time interval. The thermally generated "noise" charge is augmented by other noise charge diffusing into the depletion layer from the remaining undepleted substrate.

In contrast to the noise charges, "signal" charges can enter the CCD depletion layer by

- driving from an adjacent part of the depletion layer (i.e., normal CCD charge transfer process);
- charge packet injection from a forward biased pn-junction;

- photon absorption in the depletion layer, producing electron-hole pairs. In a p-channel substrate, holes would rapidly leave the depletion region and the electrons would remain in the potential well (the basis of the CCD television camera).

In normal CCD operation the MOS capacitor gate electrodes receive a pulsed (rather than a steady) positive bias which creates a deep potential well with instantaneous depletion width much wider than the (thermally generated inversion) equilibrium value. Electrons injected into this well, whether electrically or optically, will be stored there until two or three of the bi- or tri-phase clock pulses successively shift the potential well to the next MOS capacitor, convecting the stored charge along with it, with, hopefully, very little loss of charge.

After traversing all the MOS capacitors in a given storage block, a charge packet will have suffered a small fractional charge loss. A charge-sensitive amplifier at the output end of the storage block effectively makes up the charge loss and can reinject the regenerated charge packet into the input of the storage block.

Such CCDs have proven useful in a variety of applications including analog signal processing functions such as signal delay, general filtering (bandpass, transversal and matched) and convolution; chirp z-transform implementation; spread spectrum pulse compression; both optical and (infra red (IR) electronic imaging cameras; and line or block (vs. single word) digital memories. In this last category, however, despite their higher data storage capacity (e.g., 16 and 64 kbits) and speed (relative to moving magnetic media storage), CCDs have not outdistanced the dynamic RAM far enough in cost, capacity, or availability to succeed to any great extent. Another pessimistic aspect in this regard is the susceptibility of CCDs to naturally occurring alpha particle radiation, which exceeds even that of dynamic RAMs. Still further threat to its success as a digital memory device comes from its storage volatility and the rapid development of the high-density non-volatile magnetic bubble memory. Intel, National, Motorola and others have already ceased CCD digital memory efforts and remaining CCD suppliers such as TI and Fairchild apparently are not soliciting new application areas. A typical 64-kbit CCD memory has a 410 μ s average access time, 5 Mb/s maximum data transfer rate, and dissipates 260 mW actively and 26 mW on standby. However, IBM, Siemens, and Hitachi report 256-kbit CCD memories in research and development.

At present, most CCD development has been performed with silicon which is available with the high quality necessary to keep

charge loss (noise) below acceptable levels. Now that high quality GaAs is becoming available, much faster GaAs CCDs should be appearing before long. Already Rockwell International's Thousand Oaks Research Center has succeeded in fabricating a 259-gate buried-channel Schottley-barrier-gate GaAs CCD with charge transfer efficiency of better than 0.999. A smaller, 131-gate version with similar high transfer efficiency, has been operated at speeds up to 500 MHz. Their calculations indicate that the 259-gate version should be able to operate at up to 2.5 GHz.

Surface Acoustic Wave (SAW) Devices

In addition to the more familiar purely longitudinal (compressional) and purely transverse (shear) modes of planar pressure (acoustic) waves that can traverse the bulk of a solid material, there also exist certain combinations of these modes, called Rayleigh (surface) waves, which propagate along (i.e., parallel to) an open planar surface of the solid material. Similar to the particles of a surface water wave, particles of the solid material execute elliptical trajectories (in the plane of the surface normal and the wave propagation direction) when a continuous train of surface waves passes by. The size of the elliptical trajectories diminishes exponentially with depth into the solid material, with negligible elliptical motions much deeper than one or two wavelengths.

All of the planar surface wave power moves parallel to the surface (and not perpendicularly into the medium) so that the surface waves propagate without spreading loss. The velocity of propagation of these surface acoustic waves can be shown to depend directly on the material constants (e.g., the Young's modulus and Poisson ratio) alone and not on the frequency. Thus, to the extent that the material constants are themselves independent of frequency (up to a few thousand GHz, at least), so also, then, is the velocity of propagation, and surface acoustic waves accordingly propagate without dispersion. The wave propagation velocity is about 85-95% (depending upon the material Poisson ratio) of the bulk shear velocity and is typically of the order of a few millimeters per microsecond. At these velocities, a 30 MHz surface acoustic wave would have a wavelength of about 0.1 mm.

One practical use for such surface acoustic waves that might first spring to mind would be a non-dispersive delay line. A 10 μ s delay would require only about 2 or 3 cm of wave propagation path in dramatic contrast with the equivalent six or seven thousand feet of coaxial cable.

All that would be needed would be a thin-line source at one end, to generate planar surface acoustic waves, and a thin-line acoustic wave detector at the other end. This can be realized by plating a pair of very thin and closely spaced metallic lines on either end of the flat surface of a piezoelectric crystal (a type of crystal which develops an internal electric field when mechanically stressed and vice versa). Typical crystals used are quartz, lithium niobate and lithium tantalate.

The electrical impulse response of such a device can be estimated as follows. The spatial extent of the momentary pressure or stress distribution produced by a voltage impulse on the parallel metal strips is approximately as wide as the center-to-center strip spacing and just as long. Normally, half of the energy of the stressed region propagates away (at the surface acoustic wave velocity) toward the parallel detector strips at the other end and the other half in the opposite direction into an acoustic absorber, lest an edge-reflected "echo" follow the first half to the detector end. More advanced pressure-pulse transducer designs can generate unidirectional acoustic waves using principles akin to those of phased array antennas.

In any event, the acoustic wave pulse, generated by an extremely short electrical pulse (impulse), has a finite-width, the inter-strip spacing. Accordingly, then, it would generate at the detector a finite width voltage response pulse (even with extremely thin and closely spaced detector strips) as the finite-width acoustic pulse passes under the detector strips while moving at the acoustic wave velocity. With detector strip geometry the same as that of the generator strips, the output voltage pulse time profile (due to a voltage impulse input) is effectively the convolutional "square" of the acoustic pulse spatial profile. The impulse response time waveform has a triangular to Gaussian shape with full width at half maximum (FWHM) of about the inter-strip spacing divided by the acoustic wave velocity.

With current lithographic limits of about $1 \mu\text{m}$ resolution, an impulse response with FWHM of about $1/3$ ns results. If the impulse response is approximately Gaussian in shape, then the 3 dB upper frequency cutoff is given by $0.312/\text{FWHM}$ or about 1 GHz.

Since almost any number of intermediate detector strip-pairs can be placed between the wave-generating and final detector strip-pairs, multiply-tapped delay lines are easily implemented. If the lengths of a detector strip-pair are shortened, so that their overlap is reduced and accordingly does not intercept the full waveform of the passing acoustic wave, then a reduced electrical output results. Using this technique, called apodization, to weight

the various outputs of the tapped delay-line intermediate detector strip-pairs, transversal filters can be easily implemented.

The implementation of the transversal filter gave rise to a detector array of equally-spaced, interdigitated strips with variable lengths. Such an array is often called an interdigital transducer (IDT). IDTs can also have variable strip spacing but equal strip lengths or, in general, variable spacing and lengths. Thus, for example, if a many-strip IDT acoustic wave generator were fabricated such that in the direction going from the generator to the output detector, the IDT strip spacing and overlap both decreased linearly, then a voltage impulse to such an IDT array would result in a response function (from a very thin and closely spaced detector strip-pair) which was a chirp signal with linearly decreasing frequency and linearly increasing envelope.

Quite generally, then, an electrical impulse to an IDT array causes an acoustic wave, with the same spatial profile as the IDT array, to propagate towards the output detector. The time profile of the resulting impulse response is the same as the spatial profile of the acoustic wave incident on the detector strip-pair. Accordingly, the frequency response of the SAW device is just the Fourier transform of the IDT spatial profile function. Conversely, to implement a given frequency response, the corresponding impulse response time function is obtained by inverse Fourier transforming the given transfer function and then designing the IDT apodization profile to follow (spatially) the time profile of the impulse response. Actual SAW device designs, although based on the above, are complicated by many technical details and second order effects that make the difference between modest and superb performance.

SAW devices have made a substantial impact on wideband signal processing in areas such as the following:

- delay lines: fixed and tapped (typically 0.2 - 10.0 μ s)
- bandpass filters: 60 dB stop band rejection, 0.1dB pass band ripple; 0.2 - 50% fractional bandwidth; 1.1 shape factor (i.e., 3 dB to stopband frequency difference)
- matched filters
 - fixed: via IDT apodization pattern
 - programmable: controlled gain at each tap of tapped delay line
- high-Q resonators: e.g., Q's \sim 20,000 at 150 MHz

- ultra-narrowband filters: e.g., 70 kHz 3 dB bandwidth for 70 MHz center frequency; 40 dB stopband rejection.
- GHz "crystal" oscillators, working at fundamental frequency; high spectral purity
- pulse compression: reflective array compressors
- spread spectrum: correlators and synchronizers
- Fourier transform signal processing via chirp-transform technique: e.g., as spectrum emerges from SAW device (spectral frequency changing linearly with time), the undesirable spectral regions can be gated out at the times when they emerge, and the "filtered" result is put to another, inverse-Fourier-transforming SAW device.

SAW devices have a number of good characteristics in their favor:

- small size and weight
- ease of fabrication (after design and photolithography phase)
- high reproducibility
- no tuning or readjustment required
- long-term stability
- high reliability and ruggedness
- thermal stability
- linear phase characteristics
- independent magnitude and phase design flexibility: being a transversal filter rather than minimum-phase-shift device

SAW devices do, however, have some shortcomings: perhaps the most important one being price, if the specific device is not already offered commercially. Prototyping SAW components, as with any IC, is quite expensive. The non-recurring engineering charges required to design a high resolution photomask can run as high as \$20,000 and more. Once done, actual device costs are low, e.g., about \$1 (in huge quantities) for a TV-set LC-filter replacement.

Another SAW-device disadvantage is insertion loss (15-20 dB is not uncommon.)

In hopes of integrating high-frequency subsystems on a single chip, consideration is being given to building SAW devices on GaAs substrates. United Technologies Research Center (East Hartford, CT) is building SAW delay lines, filters and resonators on GaAs substrates for operation in the 100-200 MHz range, with hopes of eventually pushing the operational frequency to 1 GHz. They are also using GaAs substrates for voltage-tunable oscillator circuits with SAW elements (under a U.S. Army ERDC contract) and for transversal filters (under a USAF RADC contract.)

Magnetic Bubble Memories (MBM)

A magnetic bubble memory is basically an assembly of long, end-around-connected shift register arrays implemented on a thin magnetic film epitaxially deposited on a substrate crystal. The logical ones and zeros are represented by the presence or absence of magnetic "bubbles".

A ferromagnetic single crystal garnet is grown on a gadolinium-gallium-garnet (G₃) substrate in such a fashion that the direction of easy magnetization is normal to the thin magnetic film. The spontaneous magnetization of the ferromagnetic film assumes the form of intertwined serpentine domains as a result of a balance struck between several magnetic energy factors.

If the magnetic film were to be uniformly polarized -- one surface with all N-polarization, the other with all S-polarization -- there would be a substantial exterior fringe B-field of flux lines leaving one surface, looping around the film edges, and reentering the other surface. If the surface were polarized in checker-board fashion -- some squares N-polarized, adjacent ones S-polarized -- the fringing B-fields would be confined mainly to the checker-square boundaries. Since there is stored energy in magnetic fields, the second configuration would have the smaller stored energy and thus be energetically more favorable. Proceeding with this logic, however, it would seem that the lowest energy (and thus physically occurring) state would be one in which the polarization of individual atoms alternated, and no macroscopic (ferro) magnetism would be observed -- contrary to fact.

Actually, in the crystalline state, electrons interact through both electric (coulomb interaction) and magnetic (electron spin-spin interaction) terms. The direct magnetic term turns out to be quite small, and it is the coulomb interaction, in conjunction with the Pauli exclusion principle, which determines whether the ground state

will consist of parallel electron spins (ferromagnetic case) or antiparallel spins (antiferromagnetic case). In a few elements and compounds -- the ferromagnetic materials -- the lowest energy state is one of parallel electron spins. As a result, if two adjacent electron spins were to be at an angle, one should expect there to arise a "counter torque" (the Heisenberg exchange force) tending to return the spins to the minimum energy state of parallel spins. Quantum mechanics shows the interaction to be proportional to the vectorial scalar product of the two electron spin vectors -- i.e., proportional to the cosine of the angle between the spin vectors.

Thus, in the macroscopic picture, where small, alternating-polarity domains are energetically favorable, one might also expect a transition region (Bloch wall) between opposite polarity domains, in which successive electron spins would be progressively rotated, thereby making a smooth transition from the characteristic direction of one domain to the opposite direction of the adjacent domain. Militating against such an extended, smooth transition, however, is the fact that electron spins with components perpendicular to the surface normal (direction of easy magnetization) have higher energies (energy of magnetocrystalline anisotropy) than those parallel to the surface normal. Thus, widening the Bloch wall diminishes the Heisenberg exchange energy but thereby increases the magnetic anisotropy energy, and vice versa for narrowing the Bloch wall. Since the Bloch wall energy ultimately increases, whether the wall gets smaller or larger, there is some intermediate wall thickness which yields a minimum wall energy.

One final balance is struck between being all Bloch wall with no appreciable domain area (exchange and anisotropy energy dominated) and being one large domain with no need for Bloch walls (magnetic fringe-field energy dominated). This sets the average domain size, and it turns out that thin, elongated, intertwined strip regions of opposite polarity satisfy the two key requirements:

- No domain areas are very far from one of opposite polarity
- Domain areas and Bloch wall areas are comparable.

If the magnetic film is initially unmagnetized, half of the serpentine strip domains will have one polarization and half the other.

If a uniform external magnetic field (bias field) is applied normal to the magnetic film, some of the electron spins opposite the applied field direction will be forced into alignment with the field. The result is to enlarge the domains with spins parallel to the applied field at the expense of those with antiparallel spins.

As the external field is increased, the lengths of the antiparallel strip domains decrease and ultimately become comparable to their widths, at which point they assume circular boundaries due to "surface tension" of the Bloch walls. These circular cylinders of minority polarization, adrift in a sea of majority polarization, are called magnetic bubbles. If the external magnetic field strength is increased still further, the electron spins in the magnetic bubbles finally become all field aligned (i.e., the bubble shrinks into oblivion) and the film becomes completely uniformly magnetized.

Thus, by maintaining an external magnetic bias field normal to the magnetic thin film and within certain intensity margins, magnetic bubble domains can be made to exist within the magnetic film. Currently available magnetic bubble memory devices have $2-3 \mu\text{m}$ diameter bubbles and $1 \mu\text{m}$ bubble devices are not far off. From basic materials and magnetic properties, however, it would seem that the prospects of further significant reduction in bubble size are fairly slim.

To make a practical memory employing magnetic bubbles, one must be able both to confine them to fixed locations and access them there as well as to create and destroy them. Storage locations (bubble rest sites) are currently defined by depositing on top of the thin magnetic film, long closed-path storage loops (propagation patterns) consisting of small, closely-spaced and specifically-shaped patches (chevrons) of soft ferromagnetic (permalloy) material. The magnetic bubble domains, like little bar magnets, are attracted to these soft magnetic chevrons and, without further influence, stay closely attached. Since magnetic bubbles (again, like little bar magnets) repel one another, they must be kept 4 or 5 diameters apart. Otherwise, if two or more adjacent magnetic bubbles in a loop happen to be next to three or more empty bubble sites, the end bubble could be pushed off its present site onto an empty site, thereby changing the data pattern of ones and zeros. Thus, with $8-10 \mu\text{m}$ spacing between $2 \mu\text{m}$ magnetic bubbles, one million bubbles could be stored on a chip (substrate with deposited thin magnetic film) of 1 cm^2 or less.

Accessing storage locations on a storage loop is achieved by rotating the loop data (pattern of absent and present magnetic bubbles) until the content of a specific storage location is opposite a special detection point. Rotation of the storage loop data is accomplished by applying a rotating magnetic field (field access approach) which is tangential to the magnetic film surface (and thus perpendicular to the magnetic bias field that sustains the magnetic bubble domains). Each complete rotation of the tangential magnetic field induces a magnetic bubble to transfer from one chevron to the next. The rotating field is presently produced by

phased excitation of two orthogonally oriented coils wound around the bubble memory chip. Currently, field rotation rates are limited to 100-200 kHz, without, for various technical reasons, much hope for significant increase. Since the maximum chevron-to-chevron magnetic bubble domain transfer rate is determined by the maximum rotating magnetic field rate, and data transfer rates are directly related to bubble domain transfer rates, significant increases in data transfer rates are not very likely using the field access approach. The presence or absence of a magnetic bubble domain is at present most often detected (read) with a magnetoresistive loop whose resistance changes momentarily as the magnetic field of the bubble domain passes by.

Similarly to the read operation (detection of magnetic bubble domains), the write operation (creation or annihilation of bubbles) also takes place at a specific writing location from which the results are brought to storage locations by loop rotation. Two main techniques of bubble creation are direct production and seed replication. Direct production induces magnetic polarization reversal in a small area of majority-polarized material by means of a pulsed current loop. Seed replication causes an ever-present magnetic bubble domain (seed) to elongate and then split into two bubble domains, one of which remains behind as the seed.

Commercially available magnetic bubble memories exhibit several loop architectures. If storage size is not too great, one single loop may suffice. But with 100 kHz field rotation rates, the Fujitsu 74 kbit memory requires 0.74 second for a complete loop rotation, so that the average storage location access time is about 370 msec. For larger memories, and for speed in general, multiple-loop architectures are employed. For instance, a widely used Texas Instruments magnetic bubble memory of 92 kbit capacity has 144 641-bit long storage loops (actually 157, but for production yield purposes, up to 13 (redundancy) loops are allowed to be somehow deficient). To avoid having to have 144 magnetic bubble detectors and generators, a major-minor loop architecture is employed. The 144 (minor) loops each have a transfer point by which a 144-bubble parallel transfer to or from another (major) loop can take place. Thus, only a single detector and generator is needed to operate on the major loop contents. Since transfer of minor loop contents to the major loop constitutes destructive readout of the storage (minor) loops, two complete storage loop rotations are necessary for nondestructive readout: one to fetch (to the major loop) the bubble data from storage locations on the minor loops for reading, and a second to return the read bubble data (from the major loop) to their original positions on the storage (minor) loops. Even with this somewhat awkward procedure, the average access time is reduced to 4 ms.

To simplify procedures and further decrease access times, block replicate architecture supplants major-minor loop architecture. Each storage (minor) loop now has two exchange points (swap gates), one at each end. Also, at each end of the loop arrays there are half-major-loops (access rails) to or from which minor loop parallel exchanges or replications can take place. Thus, to write data into memory, data is serially written onto the input rail and parallel exchanged with appropriately positioned minor loop data. To read data from memory, minor loop data is rotated into replicate gate position and replicated onto the output rail which is serially read by the magnetoresistive bubble detector. A Texas Instruments 0.5 Mbit magnetic bubble memory, with 256 data (300 actual) 2049-bit loops, achieves an 11.2 ms average access time with a 100 kHz field rotation rate.

An operating magnetic bubble memory system consists of much more than a thin magnetic film on a chip. The access drive coils and bubble generator loop require shaping and drive amplifiers; the bubble domain detector requires a very sensitive sense amplifier; a very complicated controller is required to tend to the myriad of bookkeeping and timing details necessary to interface standard computer data-bus signals with the serial requirements of the bubble memory. In addition, most large (e.g. ≥ 0.5 Mb) memory systems provide error correction coding (ECC) capability. This both complicates the controller system and increases the number of non-data (e.g., redundancy) storage loops (for storing the ECC bits). Thus, when considering the power consumption of complete magnetic bubble memory systems, one is really comparing the consumption of the various peripheral controller systems. The same controller system, however, can control multiple magnetic bubble memory chips. For instance, eight 1-Mb memory chips can be combined by one controller to form an 8 Mbyte memory.

Designers are studying several areas that might allow boosting magnetic bubble memory chip capacities into the range of 4 to 16 Mb, ultimately, perhaps, to 200-300 Mb. Some of these areas are:

- scaling down conventional device designs limited by available lithographic resolution; magnetic bubbles must still be kept 4-5 diameters apart.
- wall-encoded lattice devices: the Bloch wall encircling the magnetic bubble domain can be given several distinguishable structural features; allows packing with only 2 magnetic bubble diameter spacing.

- Contiguous disc propagation paths: no gaps in magnetic chevrons; created by ion implantation into the magnetic garnet film; allows magnetic bubble domain diameters to be smaller than the minimum lithographic features.
- Current access methods: access fields produced by current-carrying layers deposited above the magnetic film, rather than by slower conventional orthogonally oriented coils; this technology promises 10 to 20 times higher clock speeds -- from 1 to perhaps 20 MHz -- than in present systems.

Parameters of some of the currently available magnetic bubble memory chips are given in table 3-6.

3.9.3.2 LSI Technologies with Near-Term Promise

An LSI technology with near-term promise will be considered here as one which, by approximately the middle of the 1980's, is the result of either conservatively projected advances (based on past growth performance) in a currently progressive technology, or of a well-demonstrated laboratory device technology close to being introduced commercially. Some examples of such technologies to be considered are VLSI in general, CMOS and CMOS/SOS, GaAs at the LSI level, transferred electron logic devices and Josephson junctions at the MSI level.

Very Large Scale Integration (VLSI)

Very large scale integration is not well-defined, but it is commonly agreed that several hundred thousand to a million gates per chip is the rough area. Designers, questioning how far miniaturization could be pushed, fairly well agree that physical laws draw the line at a gate length near $0.2 \mu\text{m}$, since below this point intense electric fields would cause electrical breakdown in dielectrics.

Random access memories have so far been the proving ground for VLSI and, true to form, the minimum geometries now in production are $2-3 \mu\text{m}$ (typified by Texas Instruments' 64 kb dynamic RAM and Intel's 16 kb static RAM). Channel lengths of $3-5 \mu\text{m}$ are common in the high-speed logic circuits of microprocessors. Minimum geometries of $0.5 - 1.0 \mu\text{m}$ are projected for the mid-1980's.

One rather drastic shot in the arm for VLSI technology advancement is afforded by the very high-speed Integrated Circuits (VHSIC) program of the Department of Defense (DoD). The VHSIC program will spend more than \$200 million over six years to develop

Table 3-6
Commercially Available Magnetic Bubble Memory Chips

DEVICE	MINOR LOOPS Total (Used)	BITS PER LOOP	FIELD RATE DATA RATE	AVERAGE ACCESS TIME	POWER
<u>TI:</u>					
92 kb	157 (144)	641	100 kHz 46 Kb/s	7.3 ms	0.7W
256 kb	300 (256)	1025	100 85	5.6	1.2
512 kb	300 (256)	2049	100 85	11.2	1.2
1 Mb	2x300 (512)	2049	100 171	11.2	1.2
<u>ROCKWELL:</u>					
256 kb	282 (256)	1025	100 91	4	0.8
1 Mb	572 (512)	2052	100 90	8	1.4
<u>INTEL:</u>					
1 Mb	320 (256)	4096	500 68	40	1.9
<u>NAT'L SEMI:</u>					
256 kb	282 (256)	1024	100 91	7	0.75

VLSI signal processors with several hundred times higher speed and computing power than today's LSI devices. To do so, the planned processors must also consume less power and be smaller and more reliable than current integrated circuits would allow.

The goal of this program is pilot production in 1986 of processors containing 250,000 gates, operating at clock speeds of at least 25 MHz, and performing several million to several billion operations per second. The gates would be fabricated with silicon (being a much more mature technology than GaAs) MOS or bipolar technology and have minimum dimensions of $0.5 - 0.8 \mu\text{m}$. The required speed and circuit density would be obtained both by scaling down current LSI circuits -- proportionately reducing such basic parameters as channel length, oxide thickness and supply voltage -- and by developing new types of system architecture.

Table 3-7 gives both current capabilities and mid-1980's VHSIC program goals and figure 3-21 compares VHSIC chip lithography goals with some current capabilities ($3 \mu\text{m}$ design rules). Table 3-8 gives similar commercial mid-1980's goals. Table 3-9 gives an idea of the relative promise of the various logic family technologies for achieving VLSI. Figures 3-22 and 3-23 show the yearly trend in LSI chip density increase and assumes comparable advances will continue. Since technology advancements in memories simultaneously improve both access time and density, the yearly improvement in access time is shown only for one of the earliest available sizes (1 kb) of MOS RAM in figure 3-24. In figure 3-25 a similar plot is given for bipolar RAMs of several sizes with estimated extrapolations in access time and size.

CMOS and CMOS/SOS

Several manufacturers have announced plans to strengthen their CMOS designs by reducing capacitance and aiming for speed rivaling that of bipolar technology, while simultaneously minimizing increases in power consumption.

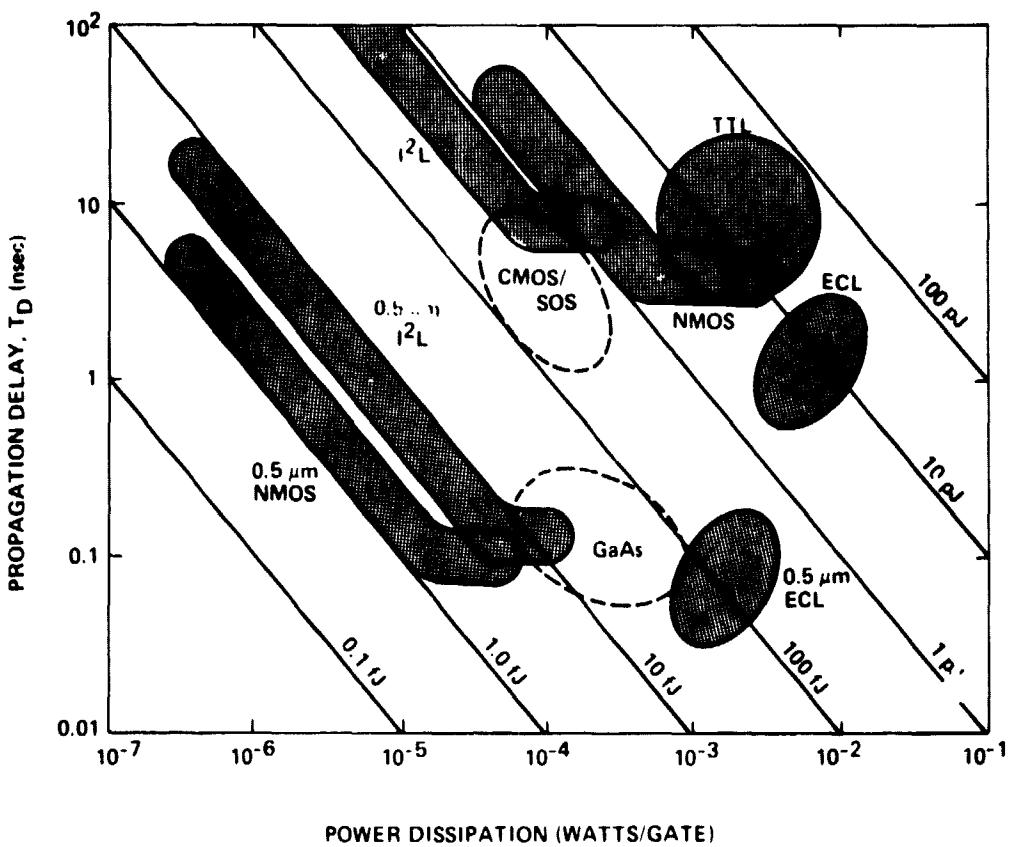
CMOS with a $5 \mu\text{m}$ gate length now exhibits a gate delay of 1.8 ns and a power-delay product of 6 pJ. The Mitel Corporation (Ottawa, Canada) is scaling their next generation CMOS down to $4 \mu\text{m}$ to operate with a 3 ns gate delay and power product of less than 0.2 pJ. Further scaling to $2 \mu\text{m}$ is expected to yield 1 ns gate delays with a 0.1 pJ power-delay product. At present Mitel isolates all n- and p-type devices with oxide to boost speed.

American Microsystems, Inc., (AMI) is scaling their next generation CMOS to $3.5 \mu\text{m}$ and expects to break the 1 ns barrier in gate delay. The company is focusing on large memories and

Table 3-7

Dod VHSIC Silicon Technology Goals

<u>Parameter</u>	<u>1979</u>		<u>Mid-1980s</u>	
	<u>MOS</u>	<u>Bipolar</u>	<u>MOS</u>	<u>Bipolar</u>
Feature Size (μM)	2.5	2.5	0.5	0.5
Gates/Chip	5 k	5 k	250 k	250 k
$T_{\text{propag. Del.}}$ (ns)	25	5	5	1
Gate Power-Delay Product (pj)	2	2	0.02	0.08
Max. Freq. (MHz)	10	50	50	250
Throughput ($F_{\text{max}} \times \text{Gates/Chip}$)	5×10^4	2.5×10^5	1.3×10^7	6.3×10^7



1A-59-561

Figure 3-21. DOD VHSIC Chip (0.5 μ m) Lithograph Goals
(Contrasted with some Current Technologies)

Table 3-8

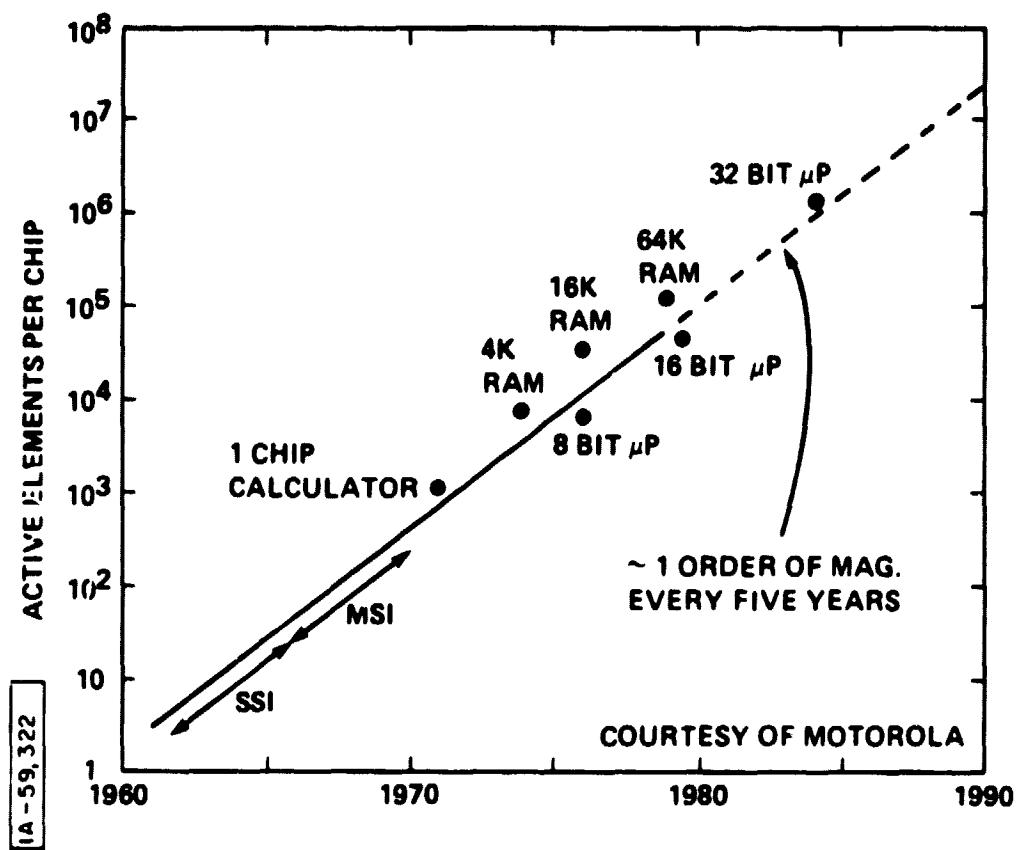
MOS: Current Capabilities and Goals

		<u>1978</u>	<u>1985</u>
Speed	Clock Rate	5 - 10 MHz	100 MHz
	Gate Delay	10 ns	1 ns
	Gates/Chip	7 - 10 k	200 k
	Mem. Bits/Chip	64 k	1 M
Density	Devices/Chip	75 k	2 M
	Dissipation/Chip	1 W	1 W
	Dissipation/Gate	100 μ W	5 μ W
	Dissipation/Mem. Bit	10 μ W	1 μ W
Power	Pinouts (Typ.)	16 - 24	24 - 40
	Pinouts (Max.)	40 - 48	64 - 120
	Clock Rate x Gate Density Product	3.5 - 10 x 10 ¹¹	2 x 10 ¹³

Table 3-9

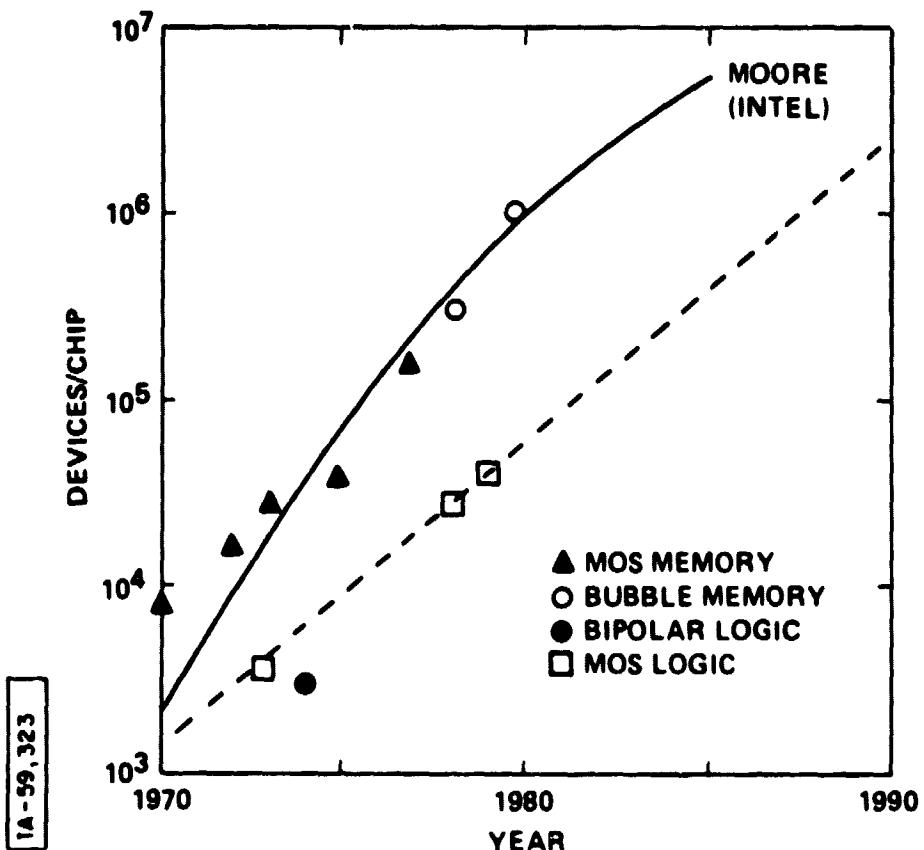
Technology Choices for VSDI		$\begin{cases} \leq 1 \text{ ns gate delays} \\ \sim 10^6 \text{ devices/chip} \end{cases}$		PROBABILITY OF IMPROVEMENT	
LOGIC FAMILY	DENSITY	SPEED	CLOCK RATE x GATE DENSITY PRODUCT	POWER	
TTL	Moderate	Moderate	7×10^9	Moderate	Low
ECL	Low	High	3×10^{10}	High	Low
I ² L	High	Moderate	5×10^{10}	Low	Moderate *
NMOS	High	Low	5×10^{10}	Low	High
CMOS	Moderate	Moderate	1.5×10^{10}	Very Low	High
GaAs	Low	High	$\sim 10^{10}$	Very Low	High (Long-Term)

* IBM just reported (p7, loc.cit.) 0.8 ns switching time, 100 μ W/gate for their I²L Technology.



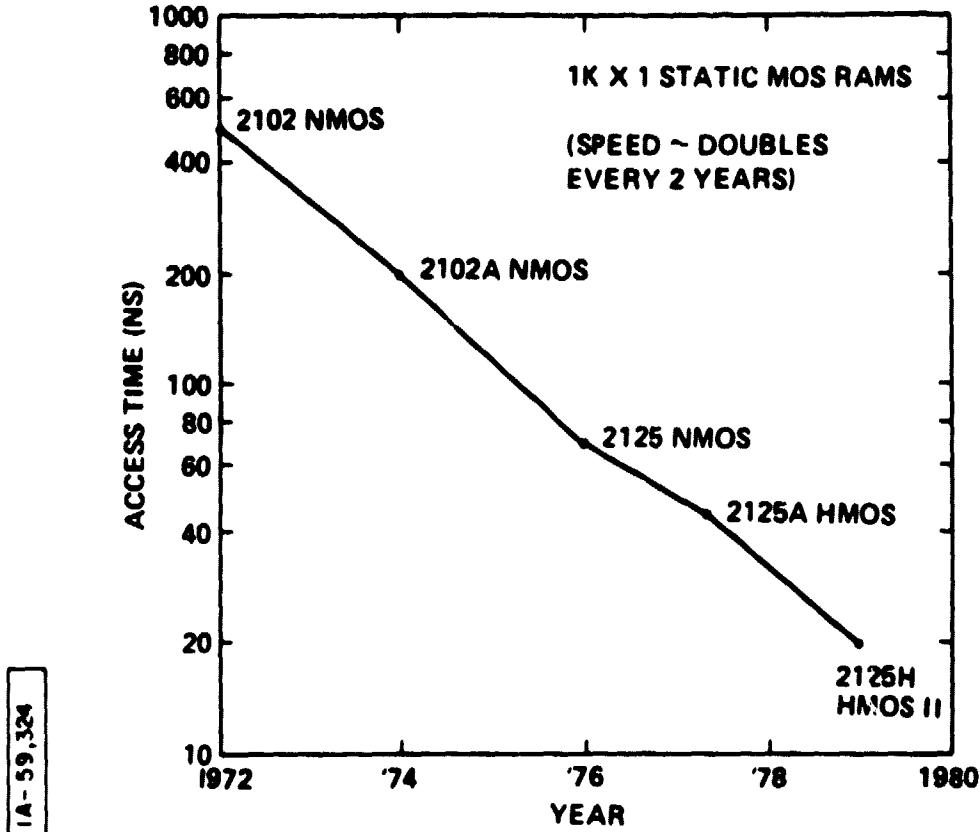
E.D. 1 JAN 4, 1980

Figure 3-22. Yearly Trend of LSI Chip Density Increase



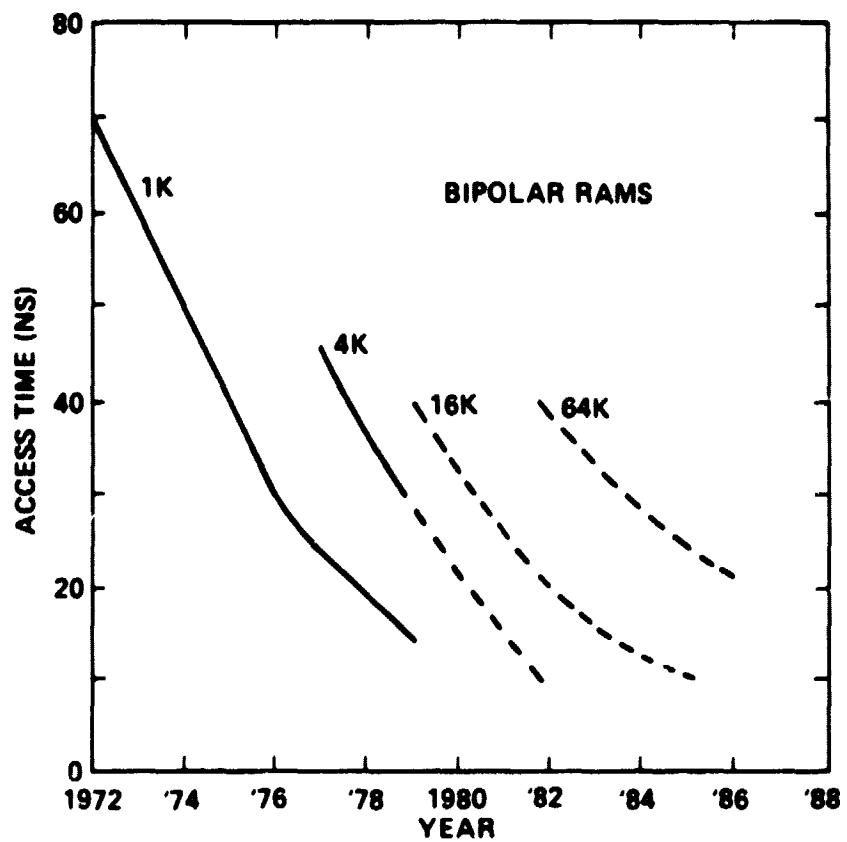
GORDON MOORE, IEEE
SPECTRUM, p 31, MAY 1979

Figure 3-23 Comparison of Yearly Growth Trends in LSI Chip Density for Logic and Memory Devices



E.D. II, 125 (1979)
ELECTRONICS, p 126, 9/13/79

Figure 3-24. Yearly Improvement in Access Time for Fixed-Size (1k) MOS RAM



E.E. TIMES, p 25
4 FEB 1980

IA-59.325

Figure 3-25. Yearly Improvement in Access Time for Several Sizes of Bipolar RAM's

microprocessors of the same complexity as are now built with NMOS, but with greater speed and decidedly lower power consumption.

Fujitsu has recently announced a 16-bit (10,000 gate) CMOS microprocessor that dissipates just 130 mW ($13\mu\text{W/gate}$) when operated at 2.5 MHz. The central processing unit (CPU) can address up to 16 Mbytes and has a 400 ns machine cycle.

National Semiconductor has introduced an 8-bit CMOS microprocessor which combines the bus structure of the Intel 8085 with the instructions and dynamic memory refresh capability of the Zilog Z80. Operating at the full Z80 clock rate of 4 MHz the microprocessor consumes just 125 mW.

Hitachi has announced a CMOS equivalent to the (industry standard 4 kb RAM) Intel 2147. Capable of being accessed in 55 ns, the Hitachi device consumes just 75 mW ($19\mu\text{W/bit}$) in the active state and only $4\mu\text{W}$ in standby (vs. the MOS 2147: 900 mW active and 150 mW in standby). They have also announced a 16 kb static CMOS RAM with 75 ns access time which dissipates 200 mW ($12.5\mu\text{W/bit}$) when active and $25\mu\text{W}$ in standby. By way of comparison, it should be mentioned that IBM has described a 16 kb static I²L RAM with 45 ns access time and 120 mW power dissipation ($7.5\mu\text{W/bit}$).

Still substantially more expensive than normal CMOS, CMOS/SOS has yet to fulfill its promise. RCA is currently using CMOS/SOS for a 4 kb memory and is planning to use the SOS capability on its forthcoming CDP-1804 single-chip microcomputer.

Rockwell International has experimented with CMOS/SOS and has developed three high-performance circuits for evaluation. One is a 12-bit analog-to-digital converter that operates in $2.5\mu\text{s}$; the second is a frequency synthesizer capable of operating at 160 MHz while consuming less than 25 mW; and the third is a Viterbi decoder that can pass a 127-bit pseudocode pattern when operating at 70 Mb/s and consuming only 40 mW. These circuits are fabricated using 3-4 μm channel lengths, which in turn yields average propagation delays of 1-2 ns and 600 fJ speed-power products. Future projections include channel length reductions to 1-2 μm , speed-power products dropping to 200 fJ, propagation delays cut to 200 ps, and packing densities of less than 1 square mil per transistor.

Table 3-10 gives some late-1980's projections for CMOS/SOS as well as for Si and GaAs.

Table 3-10

IC Device Performance Projections: 1985 - 1990
 (Lockheed Microelectronics Center)

<u>Technology:</u>	<u>Silicon</u>	<u>CMOS/SOS</u>	<u>GaAs</u>
Speed	200 ps	100 ps	25 ps
• Logic IC's	200 ps	No	1 GHz
• CCD's	250 MHz	4 GHz	40 GHz
• Transistors		8 GHz	
Optical Devices			
• FET/Laser	No (Indirect	No (Indirect	Yes (Direct Gap
Integration	Gap Transition)	Gap Transition)	Transition)
High Temperature	200°C	200°C	350°C
Operation			

Gallium Arsenide - LSI Level

Because the digital IC technology of gallium arsenide is far less developed than that of silicon, it is quite difficult to estimate the type and amount of future GaAs technological advancement. On the other hand, judging from the very impressive results obtained from the currently immature GaAs technology, a very bright future for substantial improvement seems quite likely.

GaAs will not, it seems, enjoy the great boost afforded silicon VLSI development by the DoD VHSIC program. The difficulties in GaAs IC technology are substantial. The materials technology itself is relatively immature, and flexibility in controlling all device and material properties is limited. The need to define and control some device dimensions with submicron accuracy limits the processing approaches that can be used. In particular, lack of a good native oxide for GaAs eliminates many of the self-aligned processes that are used to make short-channel silicon MOSFETs. Even the many logic approaches available have not been fully explored, so that efforts could be concentrated on just one or two of the most promising avenues.

Nevertheless, a number of GaAs IC demonstration circuits have been designed and built. Rockwell International has led the way, producing such circuits as a 64-gate 8:1 data multiplexer, a 75-gate 3x3 parallel multiplier (2 ns, 50 mW), a 96-gate 217-bit pseudorandom noise code generator, a 260-gate 5x5 parallel multiplier (4 ns, 150 mW), a 550-gate dual 4.3×10^9 bit pseudorandom noise code generator and a 1000-gate 8x8 parallel multiplier (6 ns and 300 mW). By 1984, Dr. R. L. Eden of Rockwell believes circuits with complexities approaching 10,000 gates will be feasible on a chip just 5 mm on a side. In comparison, today's microprocessors with about 10,000 gates require a chip about 7.5 mm on a side, or about 2.3 times more area. He further projects that propagation delays will be reduced to about 50 ps with $1 \mu\text{m}$ MESFET gate structures and further reductions to 20 ps will be possible with $0.25 - 0.50 \mu\text{m}$ gates. Cryogenic operation at liquid nitrogen temperatures (77°K) would ultimately reduce delays to 10 ps.

Transferred Electron Logic Devices (TELD)

Transferred electron logic devices are effectively Gunn diodes with an additional Schottky-barrier gate terminal. Many III-V semiconductors such as GaAs have conduction band structures with two minima. One band minimum is lower (1.43 eV gap) at zero momentum ($k = 0$) with high k -space curvature (low effective mass: $m_1^* \approx 0.07 m_e$) and the other band's shallower minimum is 0.3 eV higher at some non-zero momentum and has a higher effective mass:

$m_2^* \approx 1.2 m_e \approx 17 m_1^*$. If an electron in the first band (with effective mass m_1^*) is accelerated by an electric field to the point where it has gained the intraband energy difference of 0.3 eV, it can transfer to the second band where it has the much larger effective mass m_2^* .

Since no (kinetic) energy has been lost in transferring between bands, the electron's velocity must be less due to the increased effective mass. This sudden decrease in velocity due to an increasing electric field amounts to a negative mobility, μ , at that particular energy. Further increase in the electric field causes further increase (although more slowly) in the electron velocity. Since the conductivity is $\sigma = ne\mu$, where n is the density of electrons, and the current density resulting from the applied electric field is $J = \sigma E = ne\mu E$, then the decrease in J (as E increases) at the critical energy point, where the mobility drops to the lower value, represents a dynamic negative resistance effect.

The two-terminal device with this type of negative resistance region is called a Gunn diode and can be made to oscillate at up to 40-50 GHz.

Fundamentally, the exceedingly high gain-bandwidth product attainable in negative-resistance, transferred electron device operation should offer extremely high logic speeds. Due to their relatively high operating power levels and the great difficulty in getting large numbers of critical threshold devices like TELDs on one chip to have operating voltages close enough to be able to work from the same power supply, TELDs appear to be unsuitable for VLSI applications. The unique properties of TELDs can, however, offer exceptional performance in specific applications.

For example, a monolithic integrated GaAs binary phase shift keyed BPSK modulator was fabricated using only three TELDs. It handled 1.5 Gb/s data on a 5 GHz carrier with a total power dissipation of 235 mW, whereas a Si IC implementation (even of inferior performance) contained 175 active devices with a total power dissipation of 850 mW.

Josephson Junction Devices - MSI Level

Josephson junction devices are based upon the phenomenon of quantum mechanical tunneling of superconducting currents through a very thin insulating film separating two superconducting layers. The phase of the wave function of the tunneling superconducting electron pairs can be affected by the magnetic field of an adjacent control current and, in turn, the nature of the tunneling current

itself, whether it remains superconducting with no voltage drop across the junction or not.

Using such elements, it has been possible to fabricate devices with the basic logical AND, OR, invert and latch functions that have demonstrated switching times as low as 10 ps which, combined with power outputs of only a few microwatts, yield power-delay products of tens of attojoules ($1 \text{ aJ} = 10^{-18} \text{ J}$). These low powers do not reflect the power consumed by the cryogenic system maintaining the devices at the near liquid helium temperatures (4°K) needed for maintaining the superconductive state.

IBM has fabricated and tested an experimental memory model for investigating the feasibility of a 16 kb RAM with Josephson junctions. There were nearly 4500 Josephson junctions in the design which included the array, line drivers and address decoders. The storage element was a single flux-quantum cell arranged in a 2 kb array. Drivers and decoders were based on the principle of current steering in superconducting loops, which is a medium speed but low power approach.

The measured read-access time of the model was approximately 10 ns. Power dissipation for a read/write cycle time of 30 ns was about $10 \mu\text{W}$, whereas it was zero for the unselected chip. Results indicated that a 16 kb chip is feasible, and the estimated access time and power dissipation was 15 ns and $40 \mu\text{W}$, respectively.

3.9.3.3 LSI Technologies with Far-Term Promise

An LSI technology with far-term promise is considered as one which is much less certain to mature and which might, perhaps in the late 1980's or early 1990's, become generally available. Only two were found that were pertinent to the baseband processor: Josephson junction technology at an LSI level and the area of integrated optics and optical processing.

Josephson Junction Devices - LSI Level

Based on the work done so far, IBM scientists have been able to outline what the first Josephson junction computers will look like and be able to do. A range of designs has been put forward from a conservative first-generation machine to designs more representative of the real potential for superconducting electronics. The conservative design involves a mainframe computer fitting into a volume about 10 cm on a side (including CPU, cache memory, and main memory) with a 16-Mbyte (128 Mb) memory and 70 million instructions/second capacity. Less conservative projections, based on capabilities already demonstrated with experimental devices,

foreshadow super computers in a 2.5 cm cube with 64-128 Mbyte memories and capacities of a billion instructions/sec. By comparison, the largest IBM computers produced today have 6 Mbyte memories, 50-80 ns cycle-times and capacities of 3.5 million instructions/second.

Integrated Optics and Optical Processing

The basic building blocks for an integrated optical circuit have reached a state of development where complete circuits are being attempted. The performance of such basic elements as coherent light sources, optical waveguide components, optical modulators and optical detectors have been demonstrated in the laboratory to be of sufficient merit to initiate serious design efforts of functional subsystems on a chip.

One such optical subsystem receiving a great deal of attention is the integrated optic spectrum analyzer (IOSA). It consists of an injection laser whose output is conducted through a planar optical wave guide beam expander to a planar Luneburg lens which in turn collimates the laser light and uniformly irradiates the acousto-optic Bragg cell area. Exiting from the Bragg cell, the deflected light is refocused by a second planar Luneburg lens and directed to a detector array. The Bragg cell is driven by a SAW interdigital transducer, in turn driven by the signal to be spectrally analyzed.

3.9.4 Overview of Commonly Available State-of-the-Art LSI Products

Selected examples of what is felt to be indicative of the state of the art in LSI design and general functional complexity are presented. Table 3-11 lists key parameters of four of the major 16-bit microprocessors presently available, of which the TI 9900 is the oldest. Table 3-12 lists the speeds of six direct memory access (DMA) controller chips which manage fast data transfers to and from the memory without involving any appreciable interaction with the microprocessor CPU. Table 3-13 lists six vendors' approaches to implementing the NBS 56-bit key data encryption standard. Table 3-14 and figure 3-26 present some of the commercially available and due-to-become available parallel multiplier circuits. Figure 3-26 also shows the multiplier power demands made by the 10 ns sampling rate of the proposed baseband processor design. Figure 3-27 indicates how slower multipliers are pipelined to maintain the sampled data rate. Finally, table 3-15 and figure 3-28 show the state of the art, real and proposed, of the analog-to-digital conversion technology.

Table 3-11

Typical Commercially Available
16-Bit Microprocessor Chips

<u>Supplier</u>	<u>Model No.</u>	<u>GP Reg.</u>	<u>Mem. Addr'd</u>	<u>Add-Time</u>	<u>Mult.-Time</u>
TI	TMS9900	16 (RAM)	64 K	3.5 μ s	13 μ s
INTEL	8086	8	1 M	0.6 μ s	24 μ s
ZILOG	Z8000	16	8 M	2.3 μ s	18 μ s
MOTOROLA	MC68000	17	16 M	0.5 μ s	9 μ s

Table 3-12

Typical Commercially Available
DMA Controller Chips

<u>Supplier</u>	<u>Model No.</u>	<u>Speed</u>
ROCKWELL	10817	0.256 Mbytes/s
MOTOROLA	6844	1 Mbytes/s
AMD	AM 9517	2 Mbytes/s
TI	TMS 9911	2 Mbytes/s
INTEL	8257	3 Mbytes/s
ZILOG	DMA	4 Mbytes/s

Table 3-13

Typical Commercially Available
LSI Data Encryption Chips
(NBS Standard 56-Bit Encryption Key)

<u>Supplier</u>	<u>Model No.</u>	<u>Speed</u>
INTEL	8294	640 b/s
TI	9940S	4.8 kb/s
AMI	56894	{ 4.6 kb/s 11.5 kb/s
MOTOROLA	DSD	400 kb/s
WESTERN DIGITAL	2001 E/F 2002 A/B 2003	1.336 Mb/s 1.336 Mb/s (to be announced)
FAIRCHILD	9414-1,-2,-3,-4 (4-Chip Set)	13.33 Mb/s

Table 3-14

Typical Commercially Available
Fast (Accumulating) Multipliers

<u>SUPPLIER</u>	<u>MODEL NO.</u>	<u>FUNCTION</u>	<u>SIZE</u>	<u>MULT. TIME</u> (TYPICAL)	<u>POWER</u>
AMD	Am25S557	Mult'r	8x8	45 ns	1.4 W
TRW	MPY-8HJ-1		8x8	45 ns	1.0 W
TRW	MPY-12HJ	Mult'r	12x12	80 ns	2.0 W
TRW	MPY-16HJ		16x16	100 ns	3.0 W
TRW	TDC-1008J	Mult'r	8x8	70 ns	1.2 W
TRW	TDC-1009J	with	12x12	95 ns	2.5 W
TRW	TDC-1010J	Accum'r	16x16	115 ns	3.5 W

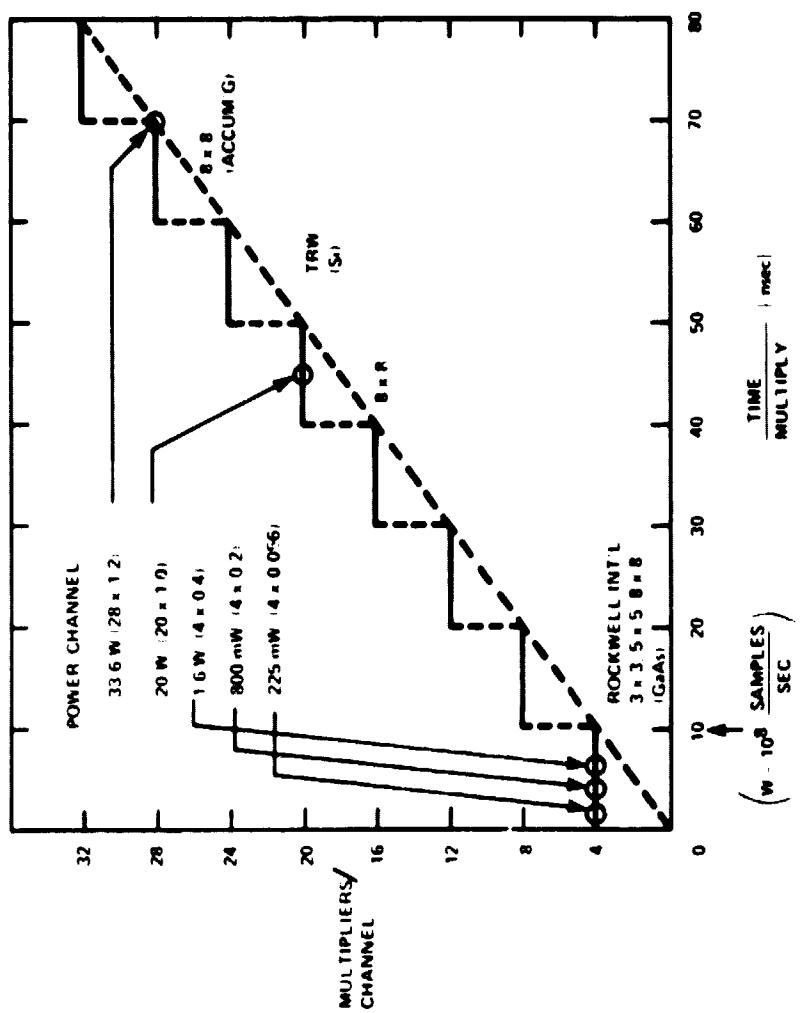


Figure 3-26. Multiplier Requirements for 100 MS/S Data Rate

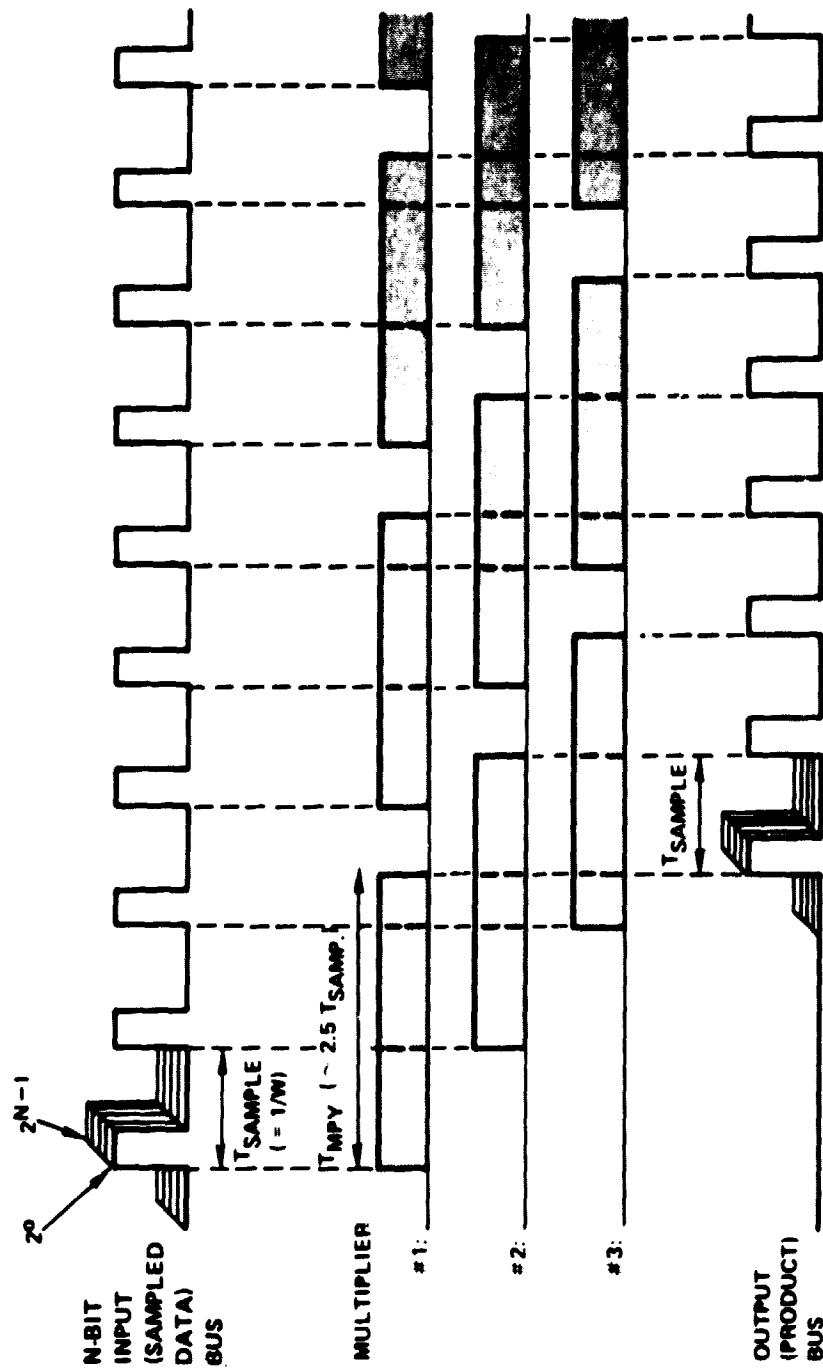


Figure 3-27. Three (Slower) Multipliers Pipelined to Maintain Sampled Data Rate

Table 3-15

**Typical Commercially Available
Fast S&H and ADC's**

<u>Supplier</u>	<u>Model No.</u>	<u>Function</u>	<u>Size</u>	<u>Sampling Rate</u>	<u>Power</u>
Datel Systems	SHM-HU	S&H		40 Ms/s	2.5W
TRW	TDC 1021J	ADC	4 b	30 Ms/s	0.25W
TRW	TDC 1014J	ADC	6 b	30 Ms/s	0.75W
TRW	TDC 1007J	ADC	8 b	30 Ms/s	2.0W

[The above ADC's, being "flash" converters, don't
require external S&H circuits]

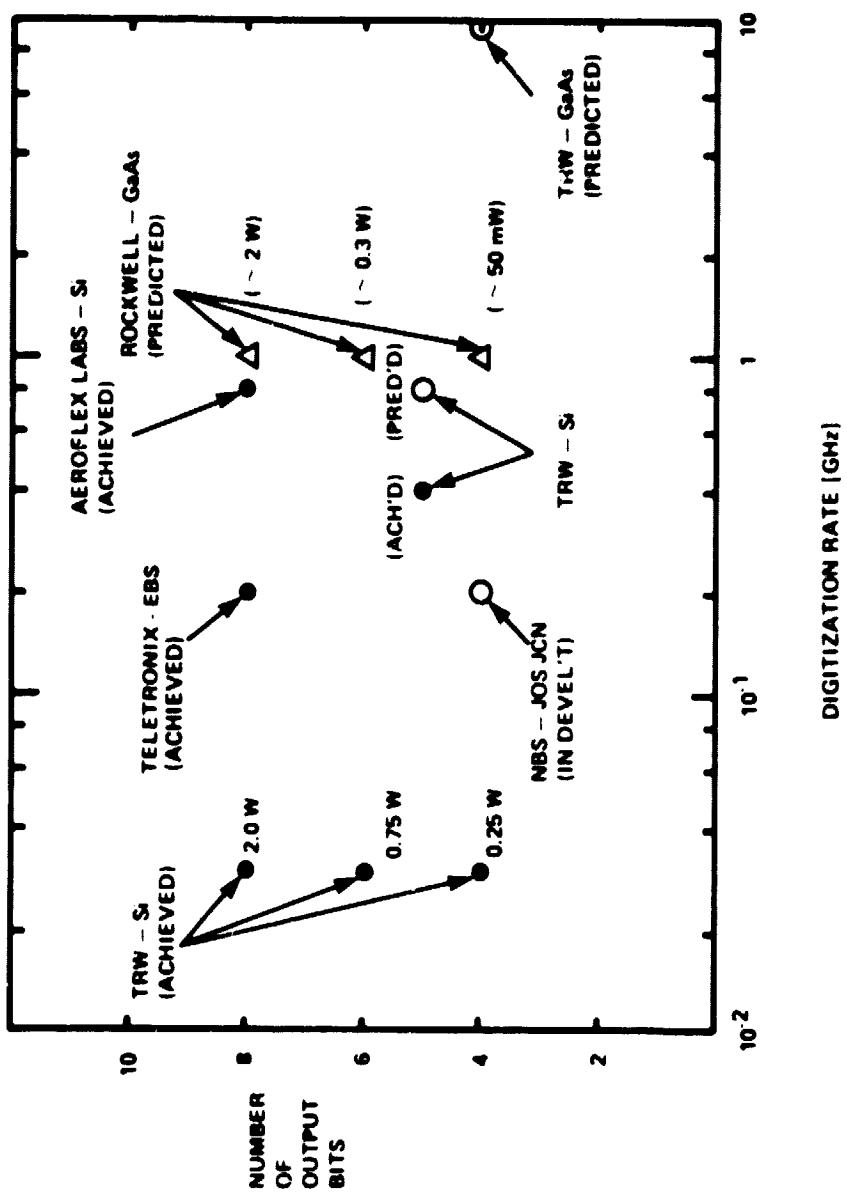


Figure 3-28. Status of A/D Conversion Development (1980)

3.9.5 Recommendations

Since the tabulation of estimated baseband processor device power dissipations indicate that CMOS memory would be relied upon heavily, radiation hardening should be looked into. With minimal vehicle shielding, an estimate of some 100 krad/yr at synchronous altitude is assumed, so at least 1 Mrad hardness is presumed necessary for a 10 year mission.

Since the speed and lower power dissipation of GaAs was also assumed in much of the processor circuitry power estimates, it is felt that it would be of great value to foster further development in the GaAs LSI technology areas of

- fast parallel multipliers,
- flash A/D converters,
- TELD modulator and demodulator chips,

as well as to look into their radiation damage and hardening situation.

More explicit information is available from the general bibliography provided following the appendixes of this report.

3.10 CONCLUSIONS/RECOMMENDATIONS

The conclusions and recommendations on the baseband processor are summarized in this section. Rough estimates of the number of IC chips and power consumption, presuming continued technological development and several customized circuit designs, are provided.

3.10.1 Baseband Processor Interfaces

In the wider sense adopted in this report, baseband processor interfaces include modulation, demodulation, and traffic parameters.

- Demodulators

The all-digital approach to on-board demodulation is feasible and is recommended pending further investigations of hybrid demodulation using acoustoelectric devices and/or charged transfer devices. The all-digital implementation would be realized either with multipliers or CORDIC rotators. For equivalent performance each technique consumes in the order of 100 W of power using fast, low power gallium arsenide MSI or LSI circuits. A more detailed comparison is necessary to strongly recommend one technique over the other.

Analysis suggests specific parameters for the all-digital design of the baseline system: (assuming 100 MHz bandwidth and 1 b/cycle modulation; see Appendix C)

8 FDMA channels per beam per dwell interval

8 b accumulators

Multiplier Approach

5 b A/D converters

5 x 5 b multipliers (pipelined)

CORDIC Rotator Approach

6 b A/D converters

4 mini-rotations per rotation (pipelined adders)

Attention should be given to the development of the required GaAs components. Further consideration and spacecraft qualification of acoustoelectric devices and is recommended (see Appendix D).

Although microwave IF (microstrip) hardware can be employed as a passive and purely analog means of demodulating QPSK, this method is not recommended because it cannot be used with advanced bandwidth and power efficient modulations which utilize a non-rectangular baseband waveform shape.

- Modulation

A constant envelope, 4-ary, or perhaps 8-ary, small modulation index, phase-coded modulation with baseband signaling shape extending over several symbol intervals should be selected for conservation of frequency spectrum and transmitter power. Emphasis should be given to discovering good coding schemes to employ as an integral part of the modulation. Attention should be devoted to the development of acceptable methods of signal synchronization and maximum likelihood detection. (See Appendix E.)

- Traffic Analysis

Further work is warranted to determine communications traffic statistics and how they can be used to bias the baseband processor controller for more efficient use of satellite capacity. Mechanisms for measuring and automatically acting upon departures from the preprogrammed average traffic loads are required. (See Appendix A).

- Non-Real-Time Data

Packet broadcast, multiple access schemes exist for low duty factor terminals with non-real-time data that make very efficient use of shared satellite channels without centralized control. Capacity and delay performance has been analyzed. (See Appendix B.) Current planning efforts should include means for incorporating such services in future satellite system designs.

- 3.10.2 Bit Processor/Packet Switch

The bit/processor or packet switch is the central portion of the on-board baseband processing activity. A general purpose architecture capable of performing a variety of functions necessary for handling customer premises traffic was postulated. The processor/switch architecture was designed in sufficient detail to permit estimates of the number and type of IC chips required. Technology assessments have led to the approximate estimates of prime power consumption listed in table 3-16. It is emphasized that considerable advanced development and custom IC chip design is required to meet these goals. The TDM double buffer appears to be very consumptive of power and therefore may benefit from some specialized implementation technique. The current power estimate is

Table 3-16
Power Estimates for Baseband Processor

HARDWARE DESCRIPTION	NO. CHIPS	W/CHIP (mW)	POWER (W)	SPEED (ns)
Demodulators (Not including synchronization hardware, etc.)			112W	
Bit/Packet Processor (Not including control hardware)			300W	
* I/O Switching Matrices (CMOS)	32	2	2x0.064	
S-P I/P-S O 32b Registers (CMOS)	160 φ	40/50	5.12/6.4	
P-P 32/12b Registers (GaAs)	160 φ	50	2x6.4	10-16
* Switch 16b Microprocessors	160 φ	500	64	
I/O 4b-16b Address Decoders (GaAs)	283	50	2x14.15	
* Bulk Memory Data Latches (GaAs)	1320 φ	50	2x52.8	
(32b-4x4b-32b)				
* Bulk Memory (2k=64x32b)	1320 φ	25	26.4	
FIFO Memory (4 12b)	1280 φ	50	51.2	300
				200
TDM Double Buffer			316	
(s-s 1544b Register) (GaAs)	2048	154	316.2	10
Control Processor			92	
FIFO Controller			8	
Bulk Memory Controller			10	
Orderwire			40	
System Control Computer			40	
* Custom Design				
* 25% Unpowered				
			Total: 826W	

based on a straightforward design using GaAs 100 μ W/b hardware not presently attainable.

If 1 kW of power for baseband processing is too much to contemplate for the next generation satellite, one could easily reduce the power requirement by simply reducing the bandwidth per beam to 50 MHz from 100 MHz, for example. This should halve the power requirement to about 500W, using a first-order approximation. Alternatively, one might wish to investigate other baseband architectural approaches.

A number of recommendations regarding the link establishment and message handling protocols, packet formats, and packet switch design of this section are listed below.

- Protocols

Separate (two level) protocols are recommended, one for link establishment and one for message (packet) transmission. These protocols should be verified and related to any standard protocol requirements before implementation. Multipoint protocols need further investigation. Additional analysis is necessary to establish time-out values.

A decision should be made regarding the location of the orderwire processor, i.e., whether it should be on-board and controlled from the ground, or on the ground. Although a ground location may be more reliable, an on-board orderwire controller is recommended primarily because of the savings in propagation delay to and from the satellite. Orderwire access schemes require further tradeoff analysis.

- Packet Formats

Further work on the details of packet error control are required, especially in terms of the efficient use of packet overhead bits.

- Packet Switch

Additional work is recommended to further reduce the chip count through the use of LSI and VLSI to conserve space, power, and connections.

The bulk memory architecture, resulting from the high throughput specification of the baseline design, is responsible for one-third the power consumption of the bit/packet processor. This may be another area that could benefit from additional design effort.

Although the listed 200 ns speed of the FIFO memories is satisfactory these chips might be reduced in size and power.

The application of queueing theory should help evaluate delay performance and buffer length requirements. Continued assessment of packet versus message versus circuit switching techniques is appropriate to identify the hybrid approach best suited to the customer premises application.

SECTION 4

MICROWAVE SWITCH

The preceding sections have been concerned with baseband processing in a satellite serving a large number of lower data rate (typically single T₁ channel) users. The 30/20 GHz program also envisions providing high capacity service for high data rate digital trunking or heavy rate service (typically T₃, T₄ channels). With extremely high capacities, high data rate demodulation/remodulation may not be feasible.

In this class of applications, switching of the undemodulated RF carriers from an uplink beam to a particular downlink beam is envisioned as the method of routing the signals in a multibeam satellite. To accomplish this signal routing, an on-board crossbar switch is needed. Conceptually, this switching will occur at an intermediate frequency, but due to the bandwidths used in the contemplated 30/20 GHz system, the frequency will be in the microwave range. Consequently, the switching unit is referred to as a crossbar microwave switch.

The size of the crossbar (NXN) will ultimately be determined by the multibeam antenna capability, connectivity, and other factors. In the first year's effort, all possible architectures and technologies for such a switch were investigated. A crossbar with directional couplers and Field Effect Transistors (FETs) as switches at the crosspoints was selected. It was concluded that a further investigation of a 100 x 100 size would reveal all of the problems which required detailed attention. (A rationale based on population statistics was also presented which suggested that a switch of this size might be useful at some future date.)

This section describes a design study of a 100 x 100 crossbar microwave switch which follows up conclusions drawn in the first year's study report.[1,3] The design is physically a microwave crossbar with FET amplifiers as switching elements.

Early in this study, it was found that a design, even to the level needed to build a working breadboard, was beyond the manpower and resources assigned to this portion of this year's program. To do that adequately might have required the addition of a mechanical engineering team, including drafting, plus augmentation of the electrical engineering team. A minimum of six man years of effort would have been needed.

Instead, salient areas were studied. Key results along with an outline of how to build the switch are presented. Present day microwave integrated circuitry based on microstrip transmission lines and discrete microwave devices is employed. These circuits are printed on high dielectric constant materials such as Epsilam 10 which reduces wavelengths by a factor of 3. The printed circuit materials are supported by epoxy fiberglass boards because the materials are only 0.025 inch thick. The material and the board are laminated and cut to a convenient size to form cards. Then the cards are employed in the card, connector, motherboard, drawer technology prevalent in current electronic equipment design. It is assumed that the printed circuit cards can have appropriate mechanical hardware which will enable them to be mounted in the kind of open framework necessary to accomplish the crossbar interconnections for the switch. The framework will have modules so that it can be assembled a drawer at a time. The drawers will be enclosed by covers which can be easily installed or removed. A conceptual discussion of the framework and card assemblage is given below but its detailed mechanical design is not considered.

Proven microwave integrated circuit technology is used in the design of the switch. The alternative was to assume the monolithic approach currently being researched for amplifiers and to adopt the predictions of the advocates regarding reductions in size. This could have led to predictions of reduced sizes for the switch which would have been highly speculative. In addition, costs would have been much higher because the entire switch might consist of a large gallium arsenide crystal. Most important, the infra structure of products needed to make the switch, such as connectors, circulators, cards, frames, printed circuit laminates, etc., does not exist for monolithic technology. Consequently, the best currently proven technology is used to configure the switch. The object of this phase of the study is to provide a benchmark design which will work and against which future advances can be measured. The sizes and weights which result are upper estimates and it is expected that any future design will be an improvement.

Construction principles can be summarized briefly. The microwave switch will consist of a control logic group and a microwave switching assembly. The control logic group will employ standard logic packages. Both the control logic and the microwave switching assembly will be mounted on cards which will be housed in drawers. These will have the necessary interfaces to the outside world including the master computer on board the satellite, the master satellite clock, and the uplink changes channel from the control station on the ground.

4.1 MICROWAVE SWITCH ASSEMBLY CARD ORGANIZATION

The microwave switching assembly will be cut into cards and housed in open drawers because of the three-dimensional interconnections required for the switch operation. If this assembly were not cut into cards, it would consist of a multilayer plane at least 8.5 feet tall, 12.5 feet wide, and possibly 6 inches thick, not including mounting support. The sectioned assembly will still perform as if it were a crossbar as suggested by figure 4-1.

The uncut crossbar switch includes two parallel layers. The top layer consists of the rows and the bottom consists of the columns. Signals enter the switch at the input to a row and proceed along it, passing the columns for which they are not destined. When the signal reaches the intended output column, it encounters a closed switch and is sent out of the row layer via a feed-through to the column layer below. The signal then exits the switch via this column. When cut, the microwave assembly yields two-layer cards.

As a first step in cutting the switch up into cards, the switch is divided into groups of adjacent rows. This can be envisaged as cutting the assembly, for example, between rows 10 and 11 and lifting off a strip which contains 10 rows plus the associated parts of the columns in the second layer below. The 10 row strip would be 10 inches high and 150 inches long. It would next be cut into 20 cards providing 1/2-inch clearance for connectors on each edge, resulting in 11 inch x 8.5 inch two-layer cards which could be mounted in drawers. Cables would have to be used between cards to maintain the row connectivity. The card interconnection concept along a row is called "snaking" after the coiling of a snake and is illustrated in figure 4-2. This shows an input connector to a drawer at the beginning of a 10 row strip. The other nine connectors are directly below it and cannot be seen. Each card has two sets of 10 connectors to provide row interconnection between cards. The cards are dual in that a column card is connected to, and just behind, each row card.

There remain the connections necessary to connect 10 strips each having 10 rows. This is done by verticalization and is illustrated in figure 4-3. The column layer of the dual card contains edge connectors which enable the columns to be reconnected to full 100 crosspoint height. Thus, the concepts of snaking and verticalization provide the topology necessary to connect the full 100 x 100 microwave crosspoint switch.

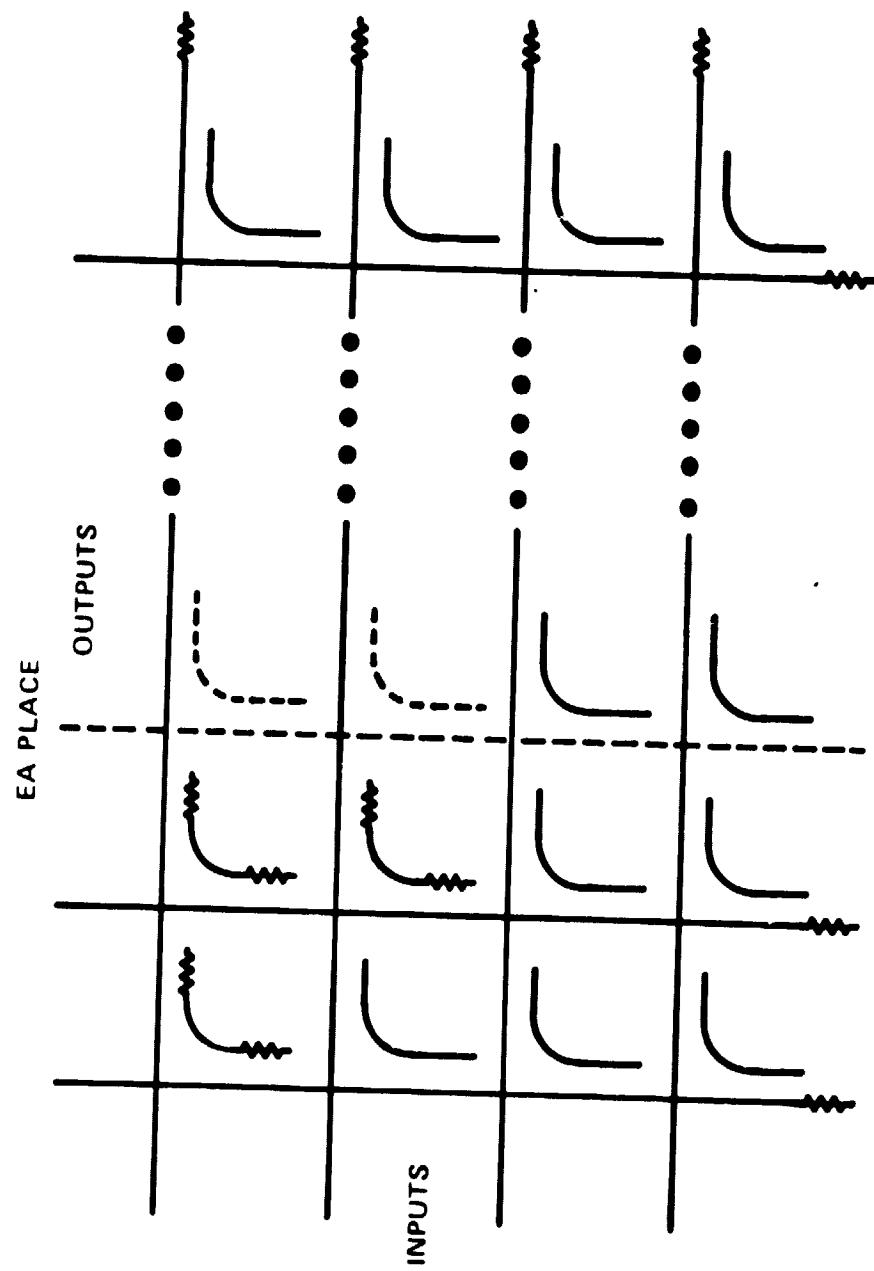


Figure 4-1. Crossbar Switch

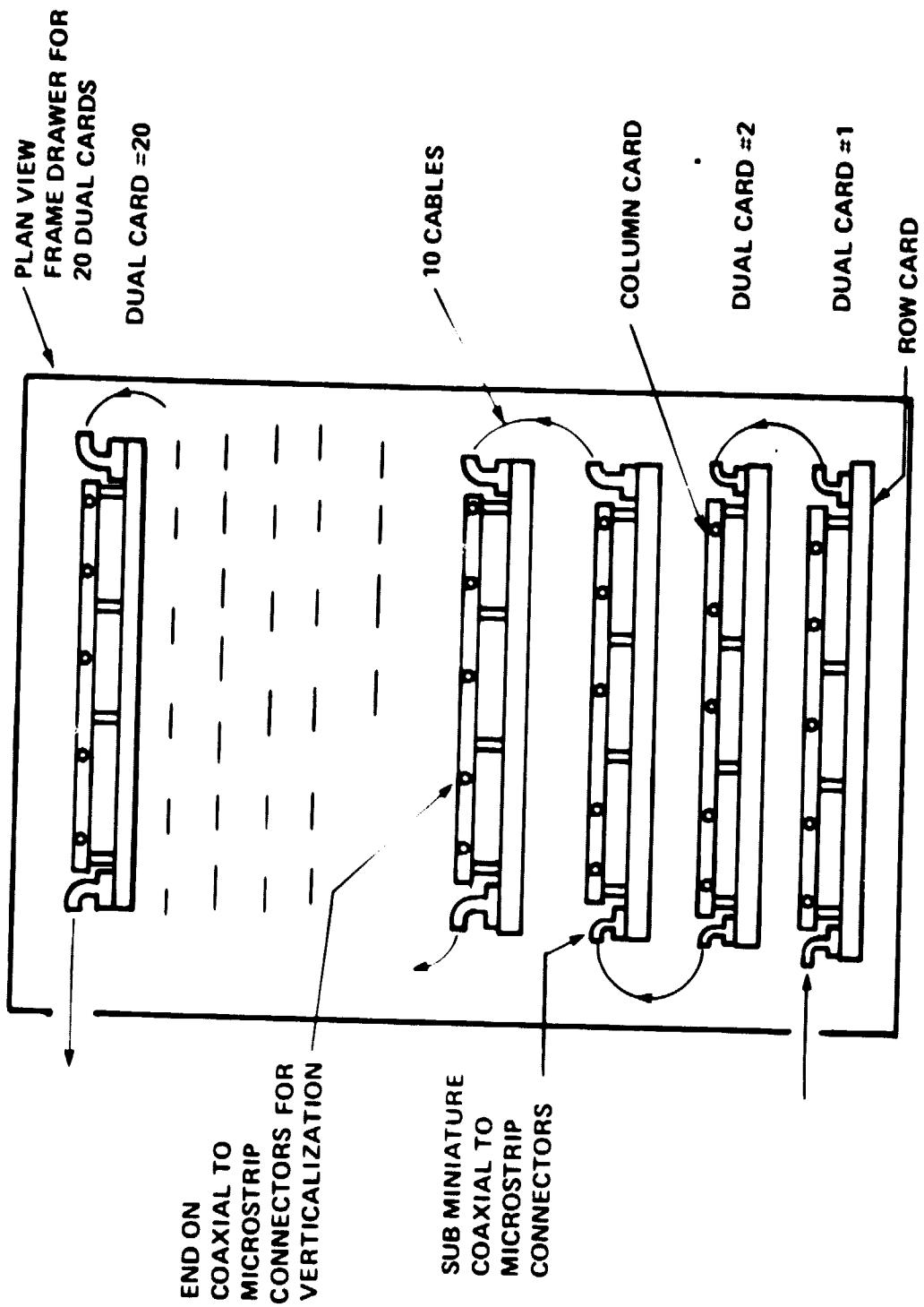
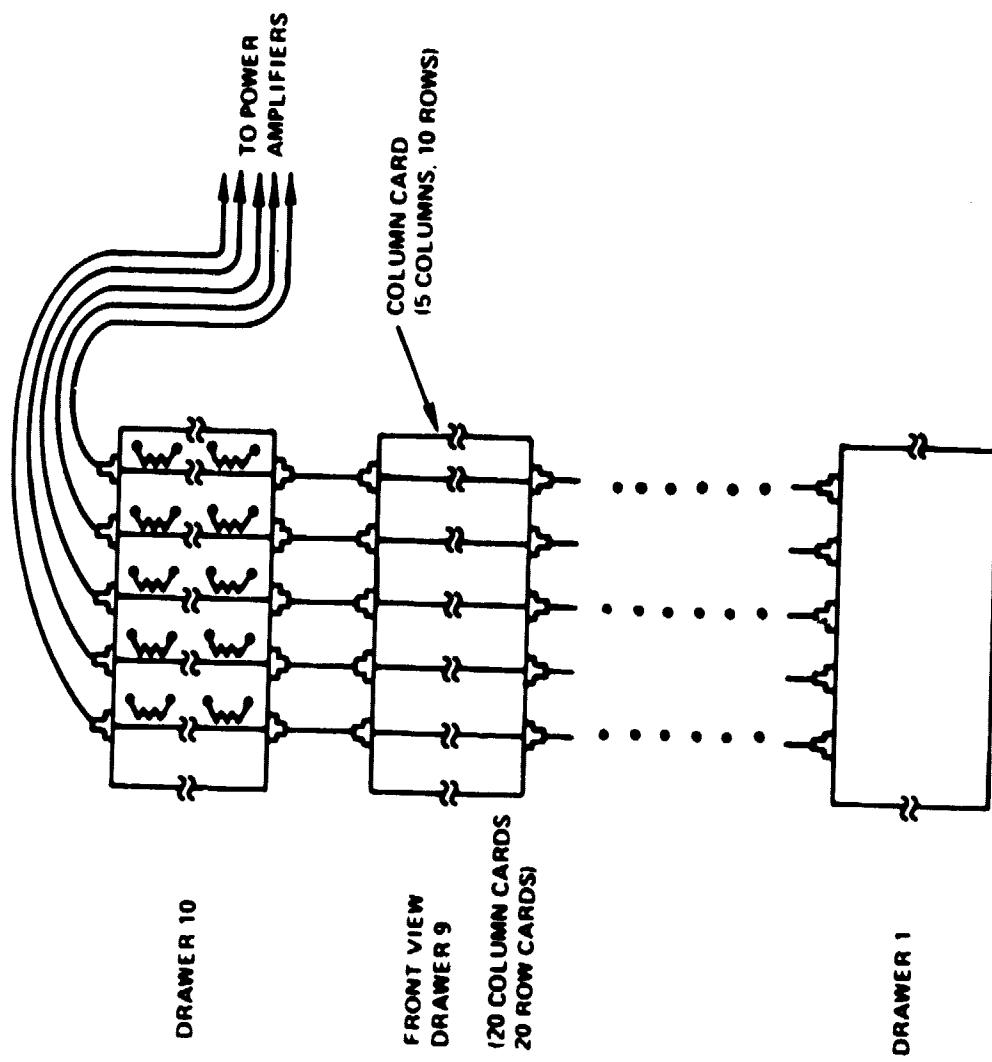


Figure 4-2. Snaking Concept for Switch Cards



The microwave portion of the switch requires 10 drawers. These drawers contain the following:

- 10,000 switchable microwave crosspoints
- 200 dual row/column microwave cards
- 100 fan out decoder cards
- 6,000 miniature coaxial connectors and cables
- 5,000 dual flip-flops for level 3 memory
- 2,000 connectors for control and direct current

The drawers must provide space for cable installation. Since the cables approach the dual cards on all four edges, a new mechanical concept is required for a drawer. This concept treats the drawer as an open frame supporting 20 dual cards but open on four sides so that necessary cabling and electrical connections can be made.

The frame can be closed with covers afterward. The concept of verticalization also requires an exterior frame so that the drawers may be supported while still interconnected top and bottom by 100 cables. Vertical spacing must be provided between the drawers by the exterior frame so that the 100 vertical interdrawer cables can be installed. It is likely that the drawers will not be stacked in one column 10 drawers high. Two five drawer stacks would be more convenient. The vertical cables can come out of the top of one stack and can go into the top of the other.

4.2 CROSSPOINT CONFIGURATION

The size of the switch is determined by that of the crosspoint. There are 10,000 crosspoints so every effort must be made to reduce crosspoint area. The crosspoint area presented is conservative. A study both experimental and theoretical is recommended to decrease its size. There is no doubt that this can be accomplished. There is also no doubt that the crosspoint described can be developed and will work. It is based upon the use of dielectric printed circuit material whose dielectric constant is 10 (such as Minnesota Mining and Manufacturing Company Epsilam 10). The dielectric material is 0.025 inch thick which requires a center conductor width of 0.044 inch for a characteristic impedance of 50 ohms.

The crosspoint shown in figure 4-4 is part of a row board. It consists of a main line microstrip conductor which ultimately passes all 100 crosspoints in a row, a wiggly line directional coupler

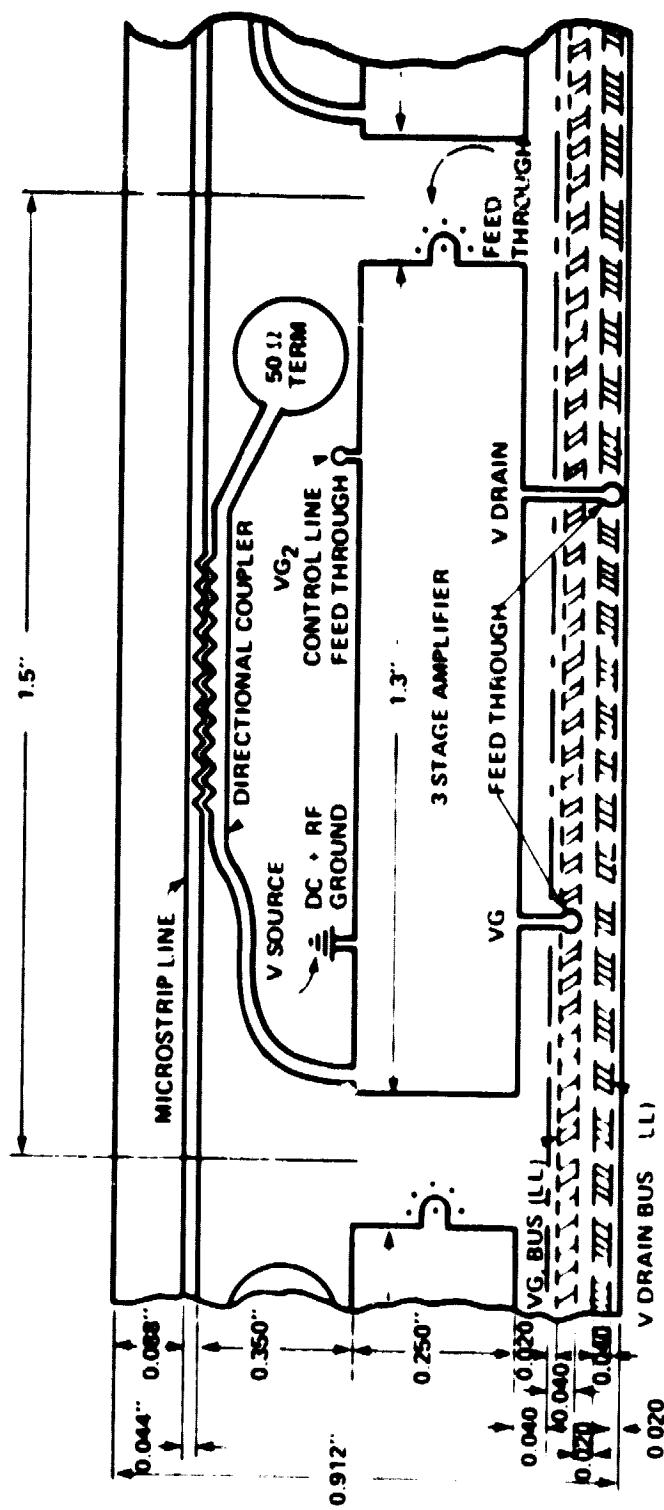


Figure 4-4. Raw Board Cross Point

which couples over a very broad frequency band, a three-stage FET amplifier used as a switch, a small matched load for the coupler, a feed-through to the column board below it, and control and direct current lines. These items require a length of 1.5 inches and a height of 1.0 inch which includes spacing between parallel rows. The signal enters the crosspoint via the microstrip line and some of it is coupled to the amplifier. If a flip-flop circuit holds the amplifier on, then the signal passes through the amplifier to a feed-through where it leaves the row board for the column board underneath.

A three-stage amplifier is used so that the control voltage, V_{G2} , need not go negative to get a 60 dB isolation through the amplifier when it is commanded open. Since the three FETs each receive V_{G2} , there are three ways to turn the amplifier off should it fail on, the more serious type of failure. Failure off is taken care of by path redundancy. The amplifier gets source and radio frequency ground connections from the printed circuit ground plane. There are common buses for the drain voltages and the number one gate voltage, V_{G1} , to which the amplifier is connected by feed-through. These lines are on the ground plane side of the printed circuit material. The switch control voltage, V_{G2} , comes from a flip-flop on the column board.

The column crosspoint shown in figure 4-5 is connected to the amplifier crosspoint by the feed-through. When the signal reaches the column board by this route, it is conducted to a second broadband wiggly coupler. The signal is coupled onto the output column line and proceeds to the transmitter amplifier for its destination city. The widths required for the output column microstrip line are so narrow that there is room available for mounting the dual flip-flops of third level memory onto the column boards. A study of the relationship of the row and column boards is given below.

4.2.1 Three-Stage Amplifier

A three-stage version of H.J. Wolkstein's dual gate MESFET amplifier will be used as the switching element for the microwave switch. Three stages are necessary to achieve the required dynamic range using a unipolar gate control voltage. As shown in figure 4-6, the amplifier stages are connected in cascade. All DC bias points (V_{G1} and V_D) have been tied together, respectively, through appropriate decoupling elements, C_1 and L_1 , to preserve system stability. The values chosen (40 pF for C_1 and 40 nH for L_1) produce a 1000 to 1 voltage divider relationship to the RF at 4 GHz, which provides adequate isolation between stages. Since this

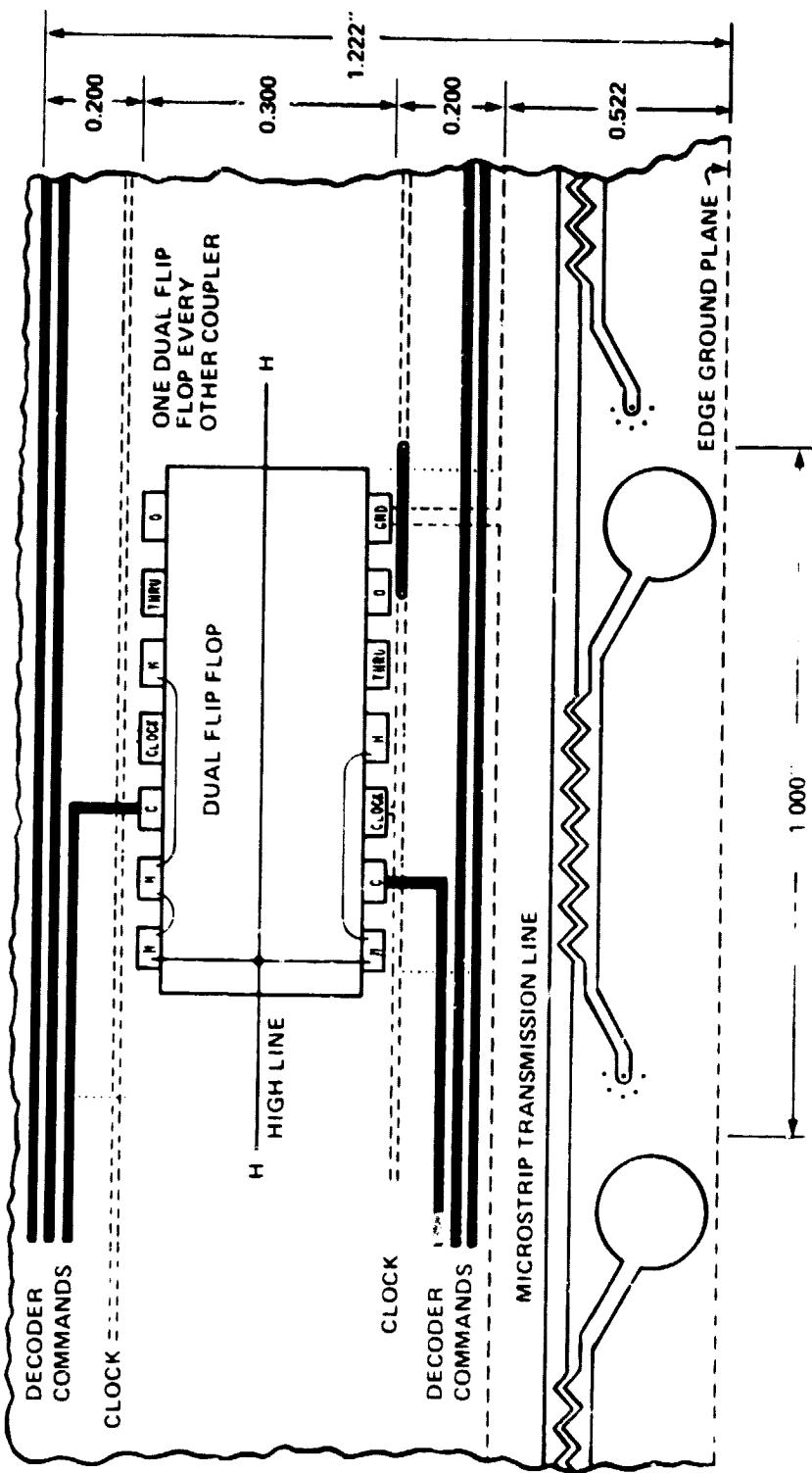


Figure 4-5. Column Board Cross Point

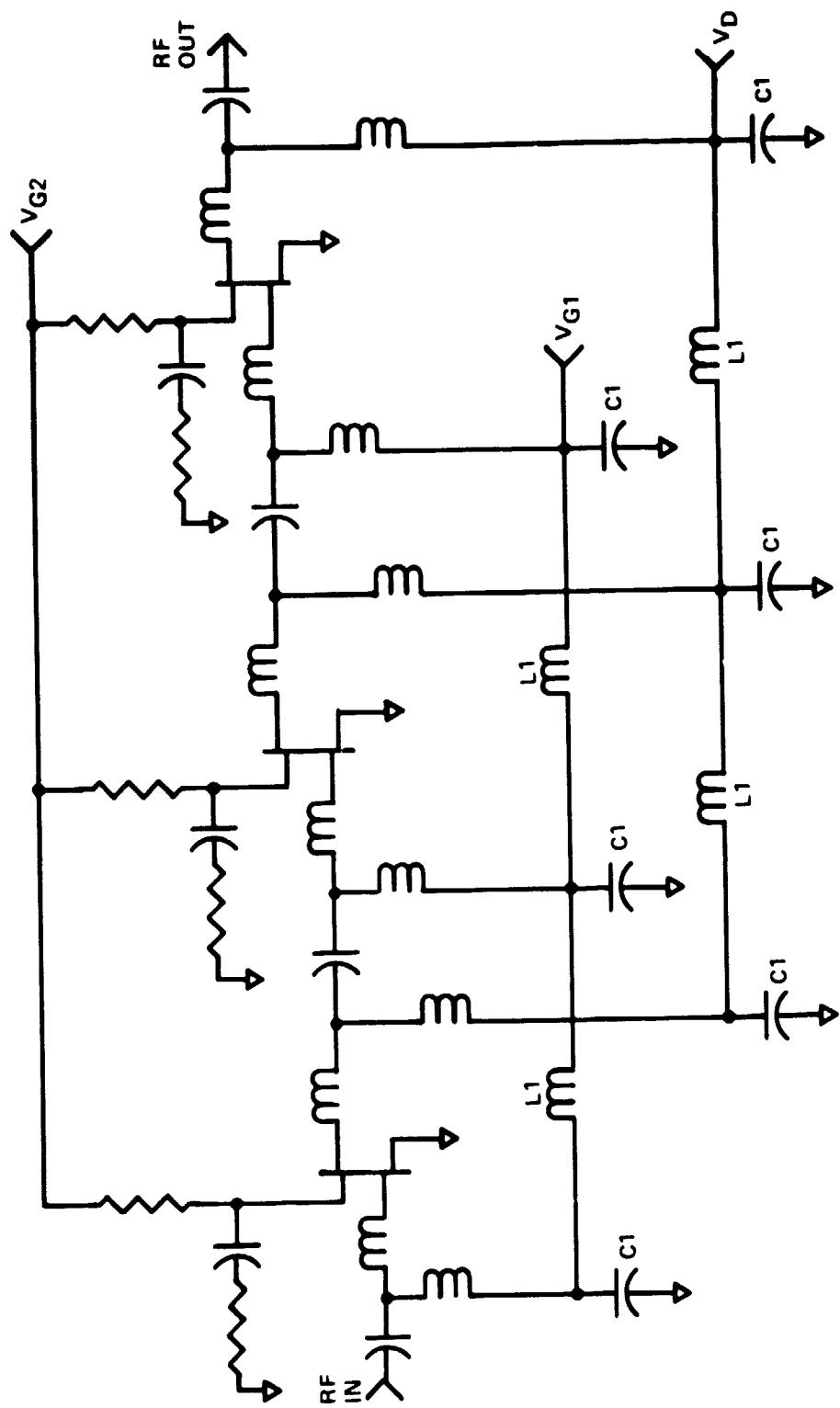


Figure 4-6. MESFET Amplifier Circuit Diagram

amplifier is being used as a switch and linearity is not of prime importance during switching, the V_{G2} control inputs of each stage can be connected. These inputs are sufficiently decoupled to render further precautions unnecessary.

Figure 4-7 shows a partial physical layout of the three-stage amplifier. The mechanical design goal is to support this circuit in an area of 1.3 inches x 0.25 inch. A mechanical analysis should be performed to determine whether the circuit should be etched directly on the PC board or added to the board as a component. Both approaches offer advantages and disadvantages which should be considered.

4.2.2 Dual Row/Column Board

Figure 4-8 shows a dual card layout study consisting of a row/column board assembly. This assembly houses a 10 x 5 crosspoint matrix and represents 1/200th of the total satellite crosspoint population. The row cards contain the input line directional couplers, the 50 ohm coupler terminations, and the three-stage FET amplifiers. The column cards consist of the output line directional couplers, their 50 ohm terminations, and the "D" type flip-flops.

Two hundred row/column board assemblies will be required to house the entire satellite crosspoint population. A significant number of RF connectors will be required. Miniature ferrodisc type circulators will be located (at some card multiple) within the vertical and horizontal RF paths to maintain a satisfactory impedance match throughout the system.

4.3 MICROWAVE SWITCH CONTROL LOGIC CONFIGURATION

The design of the control system is predicated on assumptions outlined in reference 13. They are:

1. The traffic pattern through the RF switch is statistically predetermined and encounters relatively few changes during the course of a day.
2. Only one switch in any row is active during a clock period of 1 microsecond (except for change of state). A knowledge of the 100 active switches in one clock period will totally specify the switch state for all 100 rows.

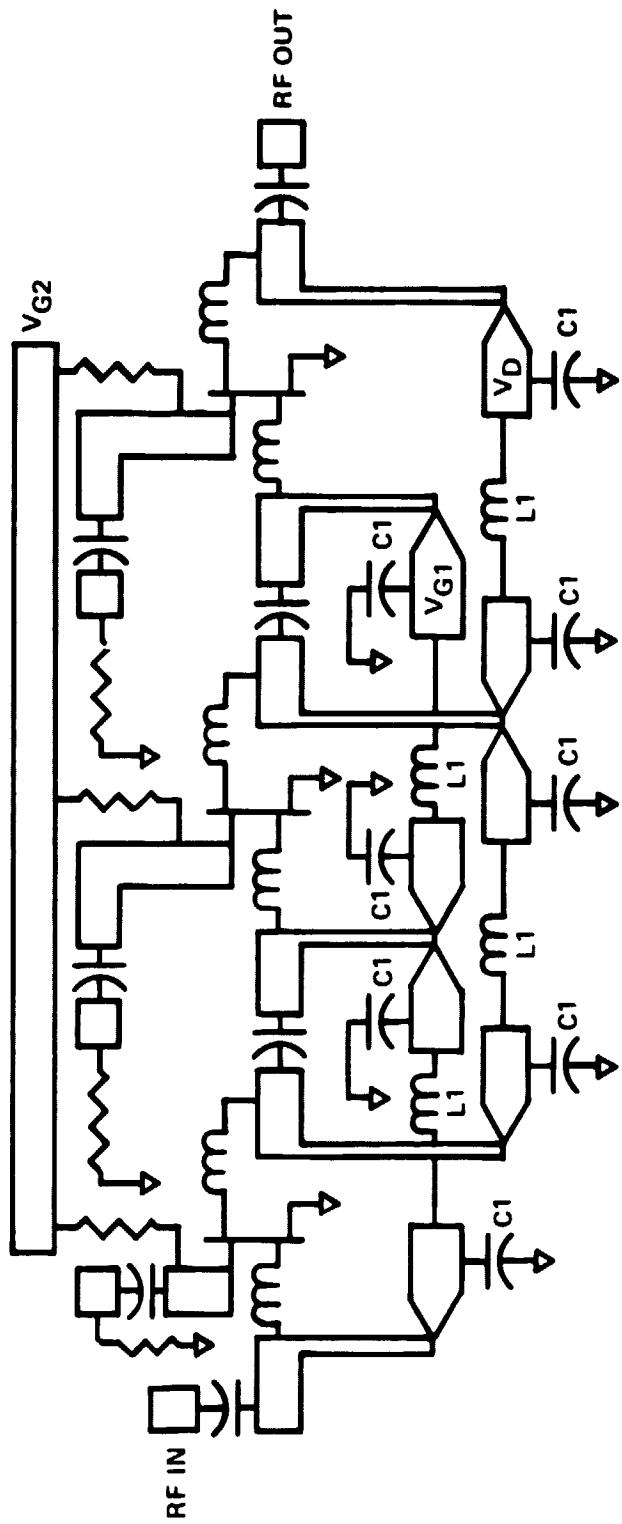


Figure 4-7. MESFET Amplifier Microwave Layout

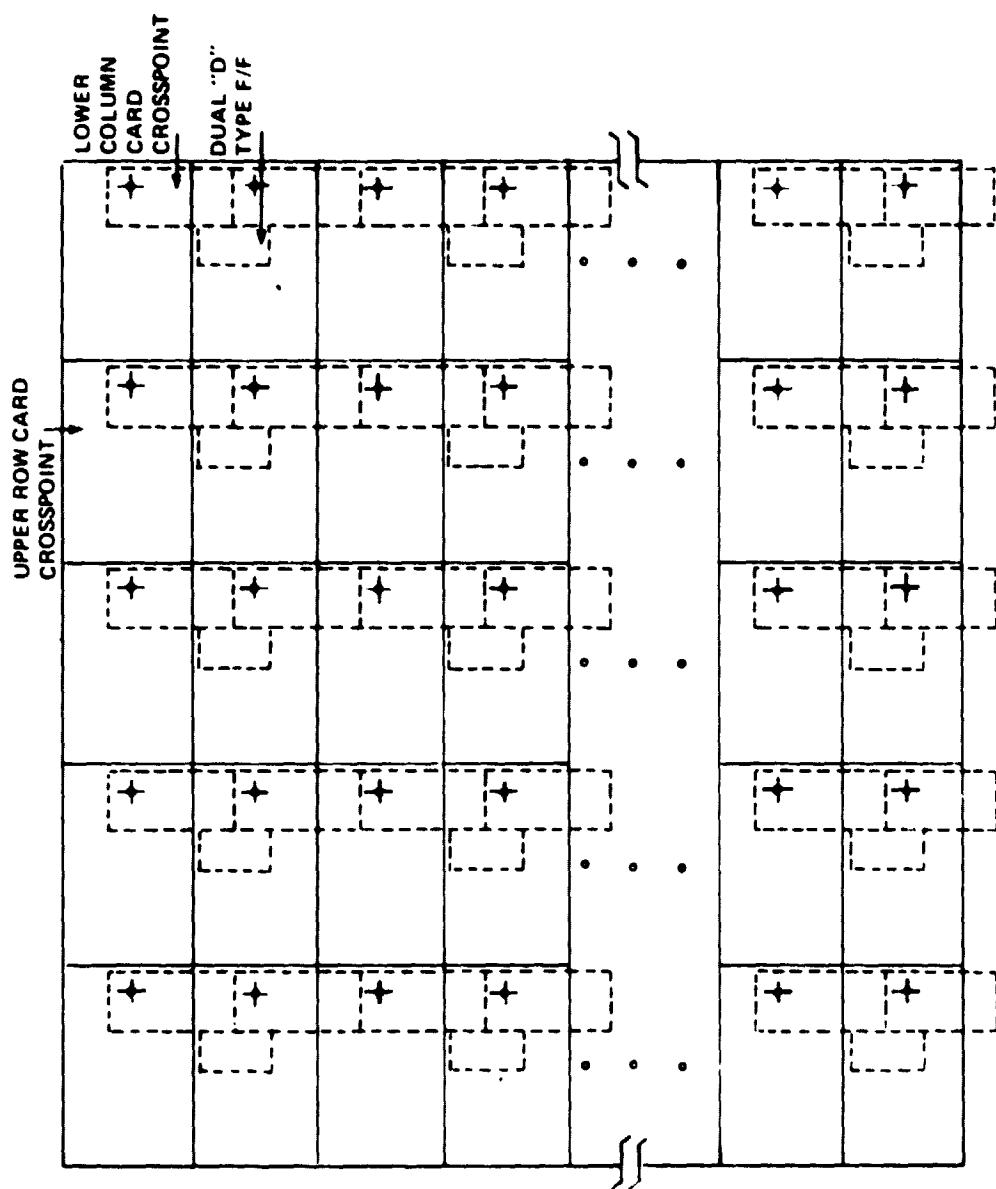


Figure 4-8. Dual Card Layout Study

3. The microwave switching sequence has a duration of 125 microseconds. This period is designated as one frame. If the switch is up-dated every microsecond, 125 states will specify switch configurations for the complete frame. The state sequence will be known as the schedule.

The schedule remains unchanged for relatively long time intervals. It will be convenient to store it in memory on board the satellite. A means of changing a portion or all of the schedule at the discretion of ground control is built into the system.

4.3.1 Memory Levels

The control system consists of three levels of memory. The third memory level is comprised of holding registers (D flip-flops) which are in direct control of FET switches and are updated every microsecond. The second level of memory is the controlling or main memory. The switch schedule is stored here. Data specifying the switch state is read out of second level memory each clock period and transferred to the appropriate FET switch holding registers. The first level of memory is used to store changes in switch states received from ground control. These changes occur infrequently. They can be transmitted to the satellite at a low data rate well in advance of implementation. The changes are transferred to main or second level memory, in response to a control command received from the ground, to update the schedule.

Only minimum memory requirements are covered in the system discussed. Ground to satellite sync., operational checks, coding redundancies, failure identification and correction, etc., are not addressed. These functions are necessary and integral to a working system and must be incorporated at some stage of system development.

4.3.2 Basic Operation

Only one switch is active during a slot in a given row of 100 switches. A 7-bit sequence specifies this switch in binary form. Specification of 100 switches in a microsecond period (a slot state) requires 700 bits. All the states of a 125 slot frame require 700×125 bits for specification. This forms the basis for level 2 memory. Ten blocks of 125 70-bit words are assumed (figure 4-9). Each memory block is connected in turn through appropriate control logic to ten 7-bit holding registers. Each register is connected to a 7×100 decoder which selects the appropriate switch in one row of the matrix (via its D flip-flop). Ten 70-bit words are read out of memory at each clock pulse (one word from each

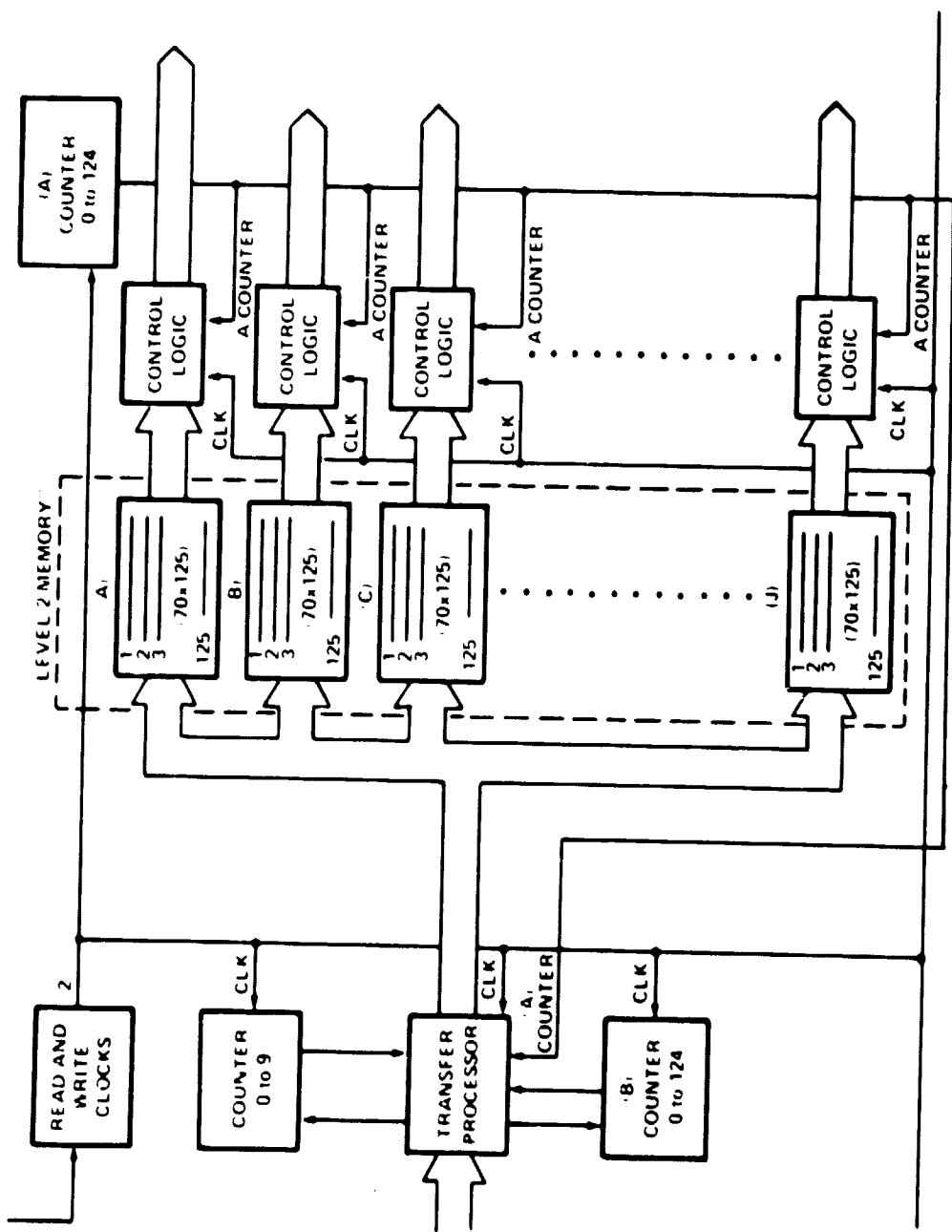


Figure 4-9. Second Level Memory Organization

block) and transferred to the array of 100 7-bit holding registers which load 100 decoders. On the next clock pulse, this information is acted upon by the appropriate D flip-flop which activates and/or deactivates its companion FET switches.

Transfer from level 2 memory to FET switch activation requires a two-clock pulse cycle. Only the time interval from D flip-flop input clock to FET switch activation is critical. This operation must be accomplished in 50 nanoseconds or less. All other transfers have a maximum of 1 microsecond for completion. No voltage translation is assumed needed between D flip-flop output and FET switch input.

A microwave FET Amplifier is used as the switch. The amplifier reported has a control voltage range of +1 to -2 volts for a dynamic range of 60 dB. Voltage translation will be avoided from positive to negative by using three stages, rather than the two stages in the circuit reported. Microwave switch reaction time of 10 nanoseconds maximum is anticipated. This leaves 40 microseconds for the D flip-flop.

Table 4-1 shows a comparison of the present logic technologies with respect to activation delay and power consumption for D flip-flop application. From this comparison, advanced low power Schottky or low power Schottky would be the most appropriate in terms of speed and power consumption for the D flip-flop/FET switch interface. Speed is not much of a factor for the remaining logic; therefore, low current CMOS might possibly be used.

Table 4-1
Comparison of Logic Technologies

Family	Max Propagation Delay	Supply Volts	Max Current Per F/F	Total Current For 10K Units
MECL	5 ns	8V	28 mA	280 A
TTL	20 ns	5V	25 mA	250 A
High Speed Schottky	17 ns	5.5V	24 mA	240 A
Low Power Schottky	40 ns	7V	3.5 mA	35 A
Advanced Low Power Schottky	13 ns	5V	2 mA	20 A
CMOS	350 ns	5V	1 mA max.	10 A

4.3.3 Level 2 Memory Update

Although the switching schedule is relatively constant over long time periods, means must still be provided for switch state update and reconfiguration via ground control. Level 1 memory and its associated circuitry provide this necessary function.

Level 1 memory is configured as 125 blocks or pages of ten 70-bit words (with appropriate address and flag bits). See figure 4-10. Each page specifies the switch state for one clock period. Set flag bit is used to inform the transfer processor that there is updated information in the level 1 memory ready to be transferred. The updated information can be loaded into level 1 memory over any convenient interval. The flag (transfer enable) bit is set when the transfer is to occur.

A transfer processor and two counters are used to transfer switching commands from level 1 to level 2 memory. The "A" counter counts continuously from 0 to 124 and recycles to 0 on reaching 124. It is used to control the readout to the flip-flops of level 2 memory. The "B" counter also counts from 0 to 124 but stops and starts to initiate transfers from level 1 to level 2 memory. The transfer processor searches level 1 memory until it detects a flag.

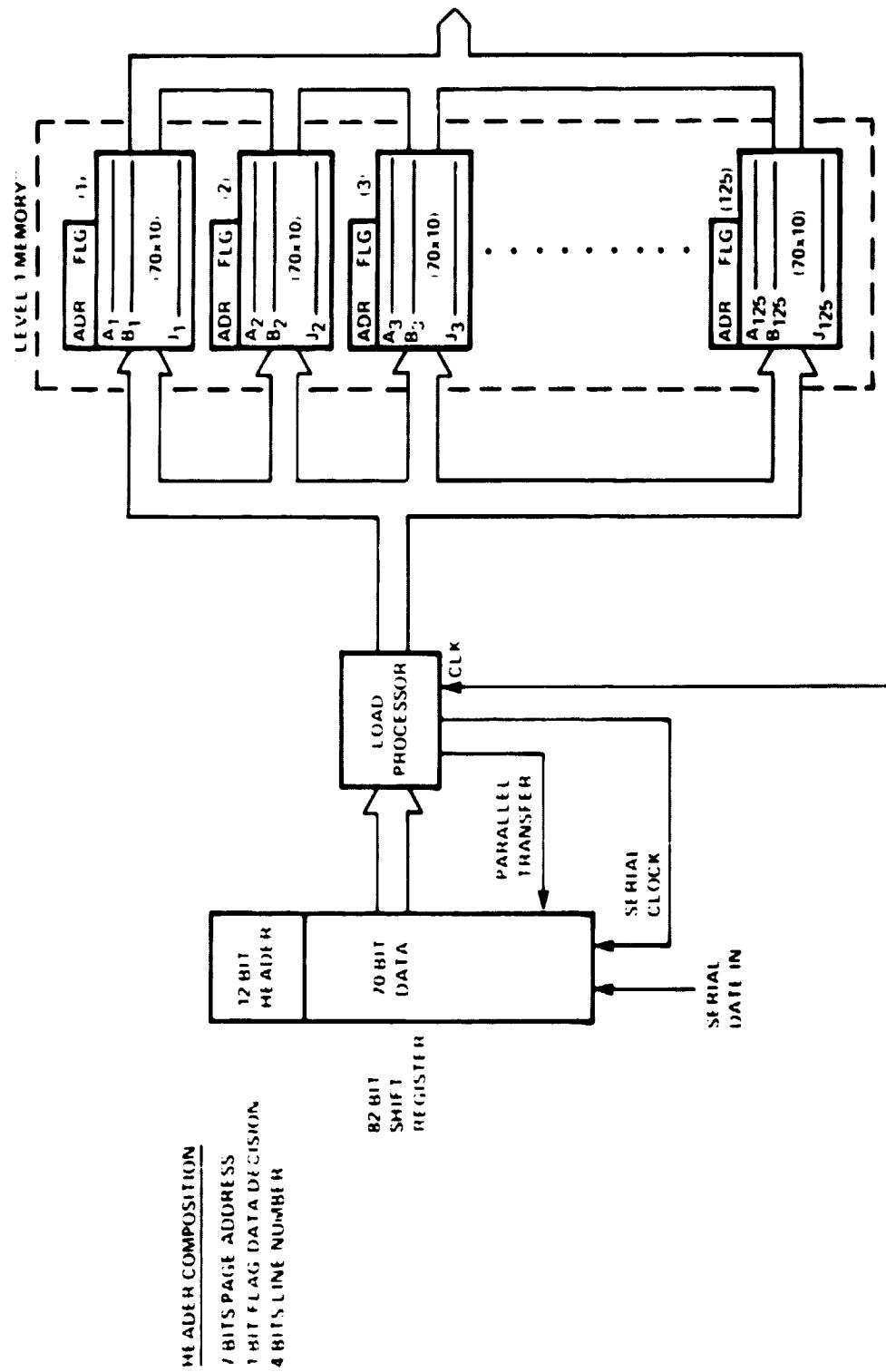


Figure 4-10. First Level Memory Organization

The accompanying "B" 0 to 124 counter is halted and a comparison is made with the running "A" 0 to 124 counter as to the current count. When counter "A" equals the contents of counter "B" plus 1, the transfer process is initiated. This comparison ensures the maximum amount of time for data transfer. The following illustration provides an example of the transfer process.

The transfer processor encounters a transfer flag in register 50 of level 1 memory. The "B" counter is halted and holds at this page number. The transfer processor monitors the "A" counter until it reaches a count of 51. At that time, the first word in the 50th page of level 1 memory is transferred to the 50th word in the first block of level 2 memory. The 0 to 9 step counter is incremented. On the next clock pulse, the second word in the 50th level 1 memory page is transferred to the 50th word in the second block of level 2 memory. Again, the 0 to 9 step counter is incremented. The transfer and increment process is continued until all ten words in the 50th page of level 1 memory have been transferred to level 2 memory. Then, the 0 to 9 step counter is set to 0, the transfer enable flag in the 50th page in level 1 memory is reset, the "B" counter is restarted, and the searching process is continued until another transfer flag is encountered.

4.3.4 Level 1 Memory Load

The switching schedule changes slowly during the day and changes tend to repeat on a day to day basis. It is possible to store changes in additional memory on board the satellite which at some predetermined time or command from ground control would be transferred to level 1 memory via the load processor. This approach forecloses random changes which might be encountered. Means must still be incorporated for direct ground update. The most flexible situation would be a combination of direct and memory update capability. Since direct ground update covers all possible reconfiguration situations, only this method is discussed in detail.

Ground update data (serial form) is read into an 82 bit shift register, whose first 12 bits are used for header (7 bits for level 1 memory page number of address, 1 bit for flag enable, 4 bits for level 1 memory line number). The load processor reads the header, determines the page number and whether data input or flag set. If flag/data bit indicates data, then the load processor reads the line number and parallel transfers the 70 bits of data from the shift register to the appropriate position in level 1 memory. Since the update period is long (hours), the level 1 memory can be loaded well in advance of update requirements. When ground control determines that switch reconfiguration should take place, it transmits a page

number and flag enable command to the satellite. The load processor reads the load command and sets the flag bit on the corresponding page of level 1 memory.

4.4 INTERNAL ELECTROMAGNETIC COMPATIBILITY

Self generated electromagnetic interference (EMI) is examined in this subsection. Cross-coupling from one control line to another must be minimized in the NASA switch. The possibility of the harmonics from the control line coupling to the RF circuitry should also be addressed.

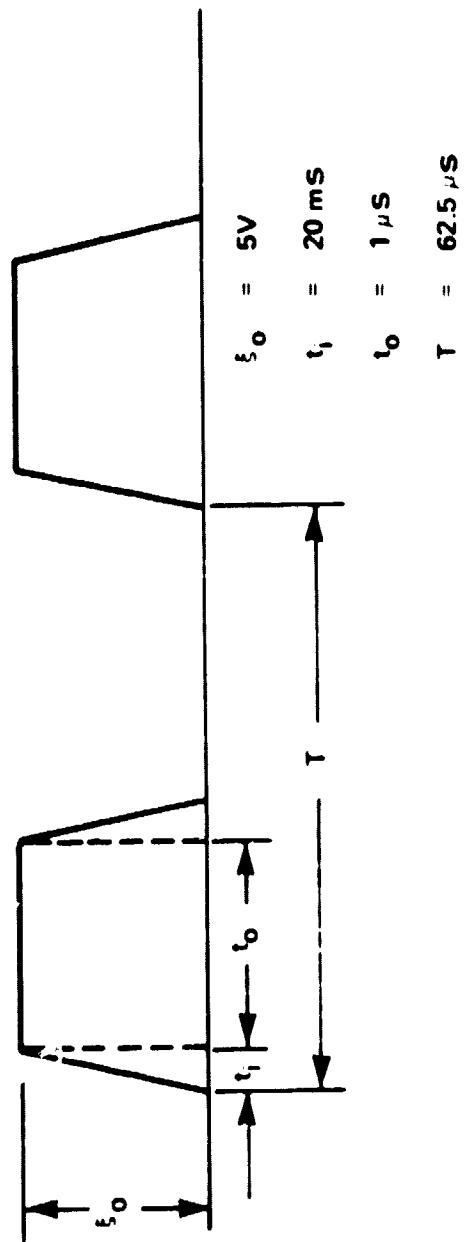
Certain assumptions are made in order to form an estimate of the level of cross-coupling and the harmonic content of the switching pulses at the signal frequency.

- The signal pulse has a rise and decay time of 20 nanoseconds and a minimum length of 1 microsecond. The highest repetition rate is 16.5 kHz. The pulse is illustrated in figure 4-11.
- The mutual impedance of the adjacent control line elements is essentially that of a parallel conductor transmission line.
- The length of the control lines are short compared to the wavelength associated with the 16.5 kHz.
- Since the control lines are electrically short, the self-impedance is that of the load, that is, the input impedance of the FET gate. The equivalent circuit of the input gate is shown in figure 4-12.

4.4.1 Harmonics of the Switching Pulse on the Signal Lines

It is conceivable that some harmonic power from the rapid rise and fall times of the relatively low frequency switching pulse train could be present and injected, by some unspecified mechanism, on the signal lines. Since this would represent a stray pulse, its presence would be unwelcome.

The calculation of the actual coupling from the switch control lines in the ground plane is a tedious task. An upper bound can be set by assuming 1:1 coupling and looking at the harmonic content of the switching pulse. If the level of any of the harmonics in or near the satellite receive band is within 20 dB of the anticipated



$$\xi_n(\omega_n t) = D_n \sin \left[2n - \frac{t}{T} \right]$$

$$D_n = A \left(\frac{t_0 + t_1}{T} \right) - \frac{\sin \frac{\pi n t_1}{T}}{\frac{\pi n t_1}{T}} \frac{\sin \left(n t_1 - \frac{t_0 + t_1}{T} \right)}{\frac{n(t_0 + t_1)}{T}}$$

Figure 4-11. Signal Pulse of Control Line

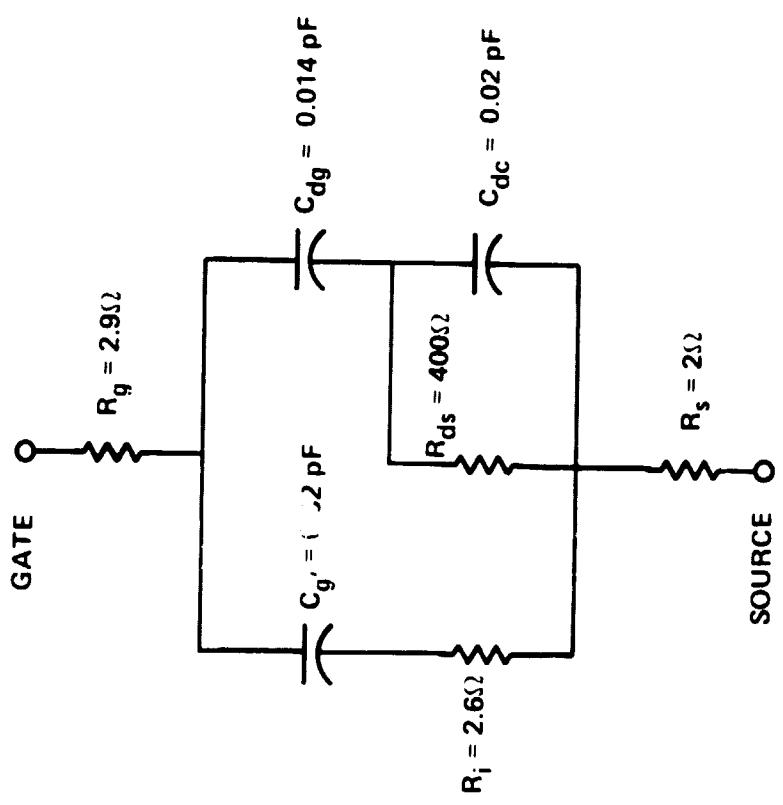


Figure 4-12. Circuit of FET Gate

signal level from the ground station, then a closer look at the actual coupling equivalent circuit is needed.

The switching pulse train is indicated in figure 4-11. The voltage attributed to any harmonic is $\xi(\omega_n t) D$ maximum. The harmonic coefficient for a trapezoidal pulse repeated at T second intervals is:[12]

$$D_n = A \frac{t_0 + t_1}{T} \cdot \frac{\sin(n\pi \frac{t_1}{T})}{n\pi \frac{t_1}{T}} \cdot \frac{\sin(n\pi \frac{t_0 + t_1}{T})}{n\pi \frac{t_0 + t_1}{T}}$$

The power contained in the nth harmonic is:

$$P_N \leq \frac{|\xi_n|^2}{Z_0} \quad P_N \leq \frac{|\xi_n|^2}{Z_0}$$

Z_0 is the impedance of the microstrip signal line.

$$\xi_n = \xi_0 \cdot D_n \sin(\omega_n t)$$

Since only limiting values are of interest, let $A = \xi_0$ (which is assumed to be 5 volts) and

$$\xi_n \leq D_n \max = \frac{\xi_0}{(\pi n)^2} \cdot \frac{T}{t_1}$$

The terms

$$\sin\left(n\pi \frac{t_1}{T}\right) \text{ and } \sin\left(n\pi \frac{t_0 + t_1}{T}\right)$$

have been assigned their limiting values of 1.

$$P_{n \text{ max}} \leq \frac{(D_{n \text{ max}})^2}{100 \text{ ohms.}}$$

The levels of $P_{n \text{ max}}$ in the band 1.75 - 4.25 GHz for harmonics falling in that band are given in table 4-2.

The three 1.25 GHz line signals will be about -40 dBW while the harmonic content from the switching pulse is of the order of -80 dB. The margin available is also shown in table 4-2.

Table 4-2
Margin of Signal Over Control Pulse Harmonics

F GHz	Harmonic Level	Signal Margin
1.75	-75.5 dBW	35.5 dB
3.00	-80.2 dBW	40.2 dB
4.25	-83.2 dBW	43.2 dB

The margins will be in excess of 35 dB assuming 1:1 coupling.

4.4.2 Cross-Coupling Switching Line-to-Switching Line

Another internal EMI likely to be present is that of a switching pulse being transferred in part to an adjacent switching control line and which causes some malfunction of the switch. Since the switch itself is basically an FET amplifier, a signal datum could partially penetrate even though the saturation level of the switch element has not been reached. The directional couplers are

taken as being about 20 dB devices and the switching FET in the off condition another 40 dB or so. If no malfunction is to occur, the line-to-line coupling threshold should be at least 20 dB below the combination, in excess of about -80 dB.

The approach to calculating cross coupling is to consider the self and mutual impedances Z_{11} , Z_{12} , place a forcing function on one line, and then determine the resulting current.

From,

$$\begin{vmatrix} v \\ 0 \end{vmatrix} = \begin{vmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{vmatrix} \begin{vmatrix} I_1 \\ I_2 \end{vmatrix}$$

$$I_2 = \frac{-Z_{21} I_1}{Z_{22}}, \quad I_1 = \frac{-VZ_{12} I_2}{Z_{11}}$$

The control gate circuit configuration of figure 4-12 indicates that the impedance of the load at the end of the switching line is:

$$Z_{gs} = \frac{(R_i - ix_{cgs})(R_{ds} - jx_{cdg})}{(R_i + R_{ds}) - j(x_{cgs} + x_{cdg})} + R_s + R_g$$

The line is electrically short compared to the wavelength associated with the switching frequency and the self-impedance terms, Z_{11} and Z_{12} , are essentially that of the load itself. With the values shown in figure 4-12,

$$Z_{gs} = Z_{11} = Z_{22} = 8.18 - j1.52 \times 10^7 \text{ ohms.}$$

The imaginary part of the impedance dominates, and it was used for the self-impedance. The line-to-line impedance of 250 ohms was used for the mutual impedances, $Z_{12} = Z_{21} = 250$ ohms.

$$I_1 = -3.289 \times 10^{-7} \text{ amp}$$

$$I_2 = -5.41 \times 10^{-12} \text{ amp}$$

The power cross-coupled is proportional to (I_2/I_1) or, P---95.7 dB. Since this implies a margin of 36 dB or more, the system, as described, does not appear to be susceptible to this form of internal EMI.

4.4.3 Signal to Switching Line Cross-Coupling

The inverse of the control to signal line coupling considers the case where 3 GHz signals cross-couple into the control line. The signal lines are assumed to be carrying about 1 milliwatt of signal power. The impedance of these lines is about 50 ohms referred to the ground plane.

The signal is assumed to be tightly coupled to the 250 ohm balanced control lines (7.3 pF capacitor). The control output voltage is the same as the control input, about 5 volts. Since nothing is known about the 3 GHz characteristics of the control load connected, a simple resistive match to the 250 ohm line was assumed.

The margin on the basis of these conservative assumptions is:

$$M = 10 \log_{10} \left| \frac{E_3}{E_0} \right|^2$$

The model which includes input and output networks as well as the FET equivalent circuit is shown in figure 4-13.

By means of Thevinin's theorem, the current in the load, R, from 1/2 V RF (1 mW on a 50 ohm line) was calculated. The margins (referred to the 5 volt switching threshold) for 1.75, 3, and 4.24 GHz are:

<u>F GHz</u>	<u>E₃ Volts</u>	<u>Margin</u>
1.75	0.0104	53.7 dB
3.00	0.0145	50.7 dB
4.25	0.0224	47.0 dB

The system, as configured, does not appear to be susceptible to internally generated EMI from the switching pulses either on the signal lines or on adjacent switching lines.

4.5 MECHANICAL DESIGN

The memory used does not have simultaneous read and write capability. To overcome this difficulty, separate read and write clocks will be generated from a 2 MHz system clock (figure 4-14). With this circuit, the 1 MHz update rate will be maintained as specified; however, circuit reaction time will be reduced to 0.5 microsecond. Analysis of the circuitry shows that this reduction can be accommodated.

4.5.1 Hardware Implementation

Level 2 memory will be configured as shown in figure 4-15. It consists of 90 CDP1823 RAM ICs, each containing 128 8-bit words. Associated with each 128 x 8 block of memory will be two CDP1856 data bus separators for input/output isolation on the bi-directional memory data bus (figure 4-16). Two additional CDP1856s will be used for the read/write address bus. Buffers will be required for the

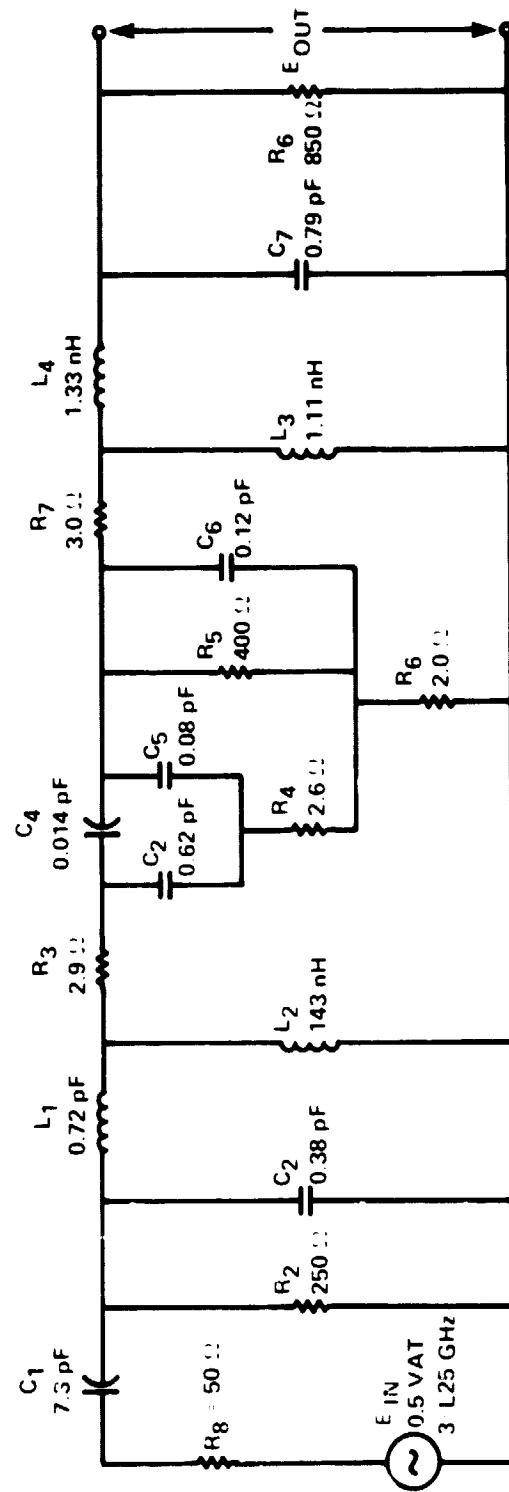


Figure 4-13. Model for Signal to Switch Coupling

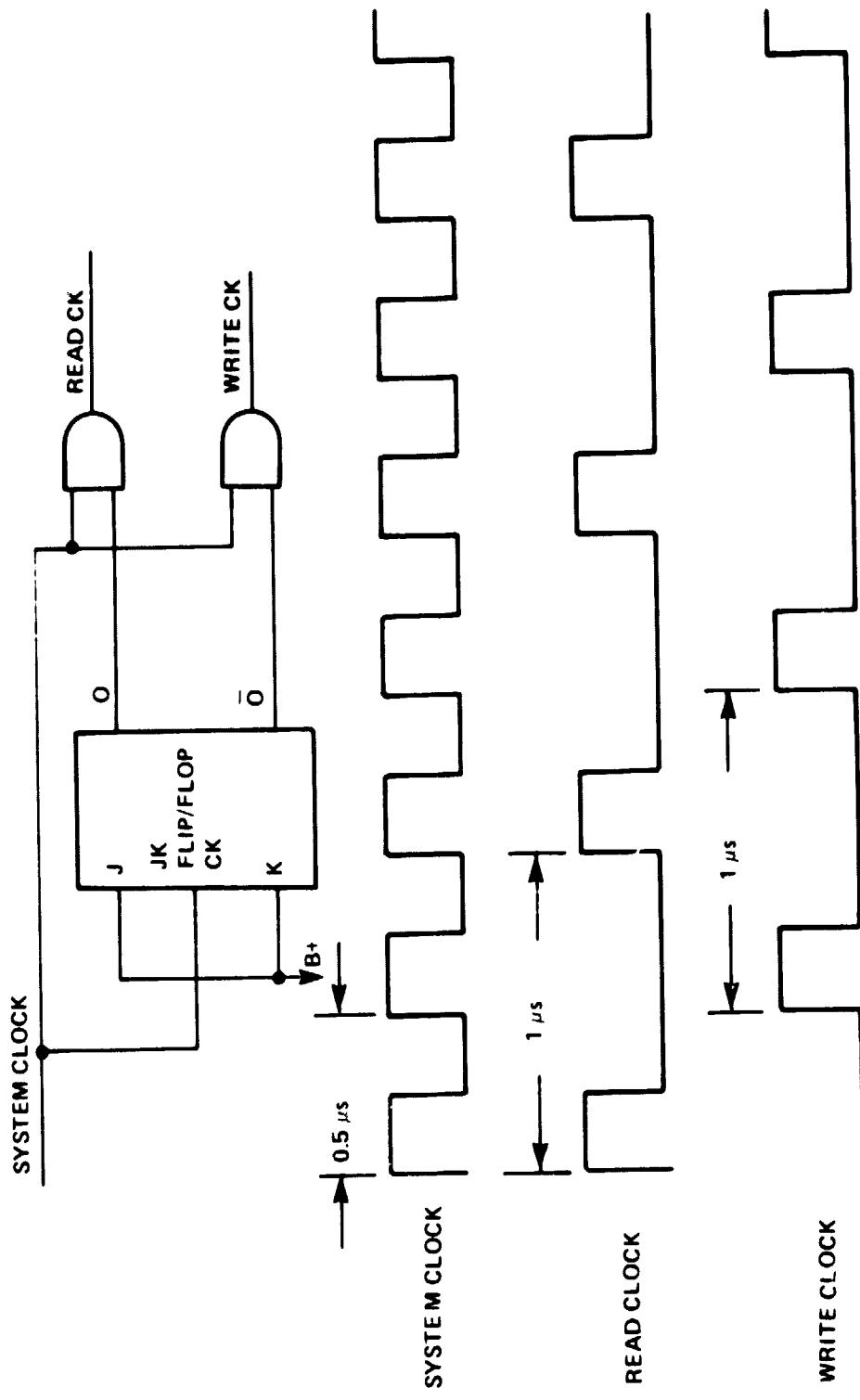


Figure 4-14. System Clock

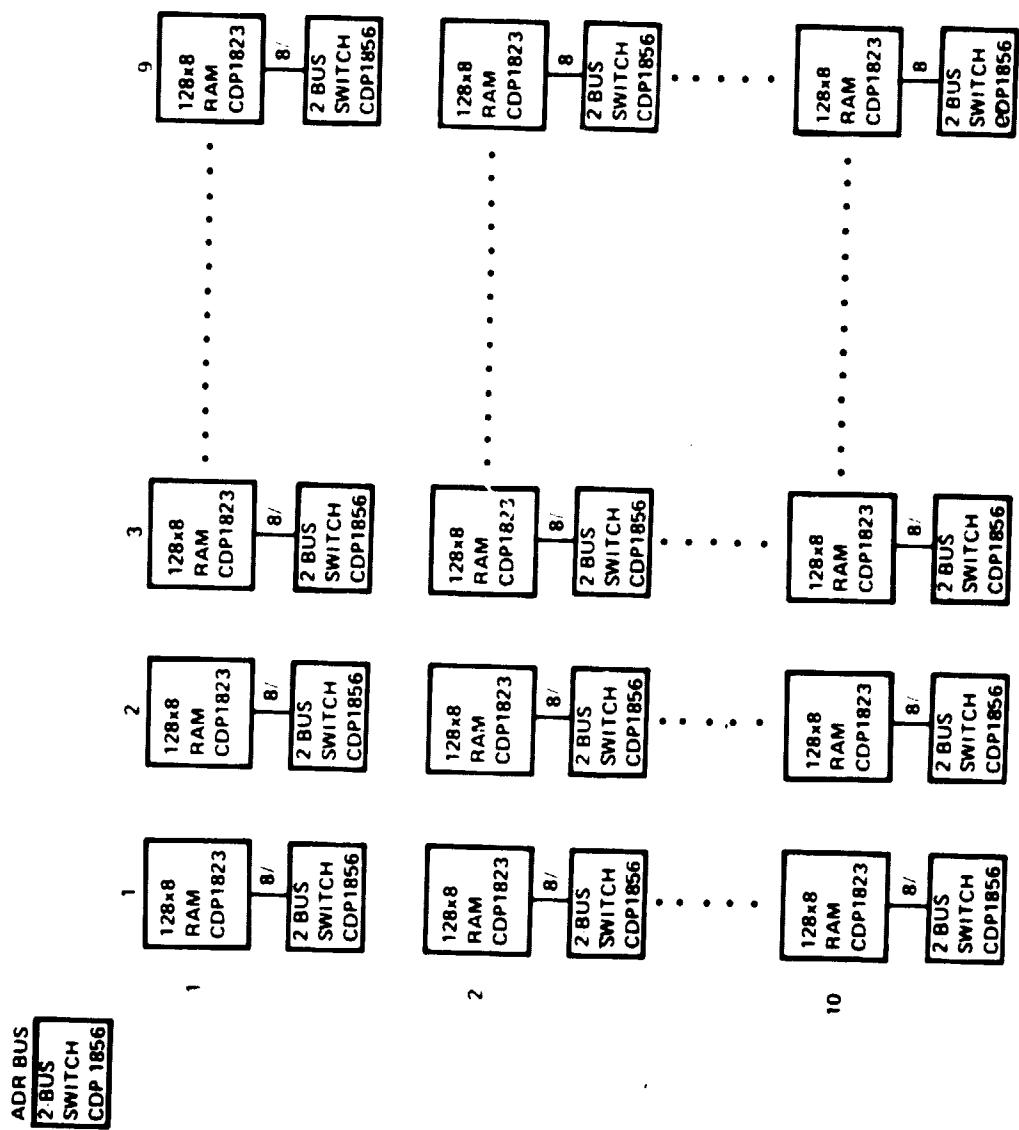


Figure 4-15. Level 2 Memory Configuration

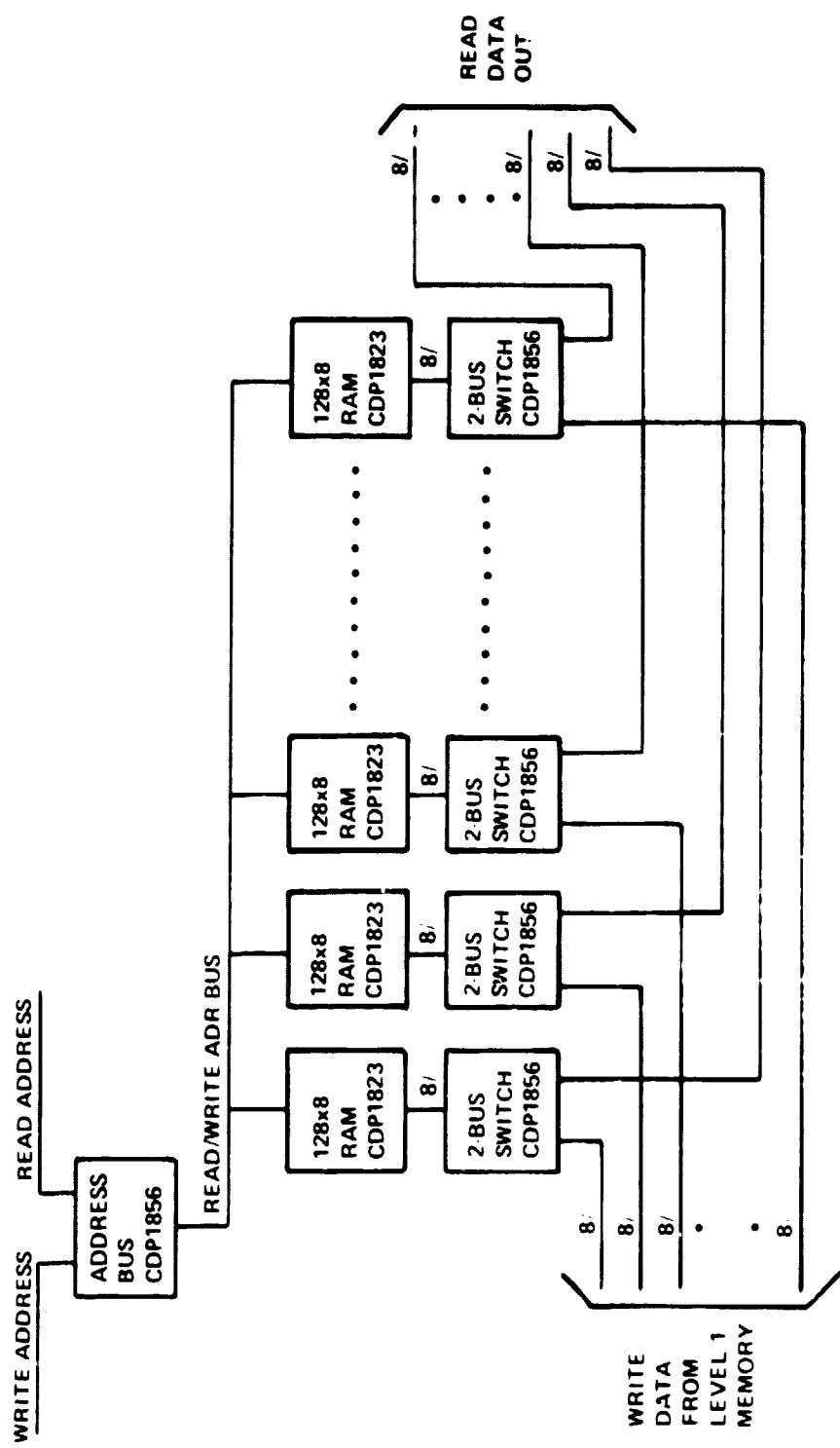


Figure 4-16. Bi-directional Memory Bus Separations

address lines of each memory IC (not shown). 110 MC14050B hex buffer ICs will be used for this application.

The level 3 memory and its associated control logic is configured as shown in figure 4-17. This diagram represents the traffic flow for one row of memory and requires a multiplication by 10 for the complete hardware count. The output from level 2 memory feeds 90 CDP1852 holding registers. These registers input 100 7 x 100 decoders. The decoders include a total of 1,500 3 x 8 CDP1853 decoder ICs. Each 7 x 100 decoder selects one of 100 D flip-flops. Five thousand SN54ALS74 dual D flip-flop ICs will be used for this application. These flip-flops are in close proximity to the actual microwave switching elements because of the rapid transfer time required.

Level 1 memory hardware is identical to level 2 memory (figure 4-15). It will consist of 90 CDP1823 RAM ICs, 182 CDP1856 bus separators, and 80 MC14050B hex buffers.

The functions of the load and transfer processors can be combined into one unit that will be designated "input processor" (figure 4-10). This processor will be the RCA CDP1804 type with internal RAM, ROM and timer/counter. The serial input registers will be comprised of 11 AM54LS299 type, 8-bit universal shift registers.

Table 4-3 summarizes the preceding information and outlines the sizing and power requirements for this control system. Using present technology and the components specified, a total of 7,385 individual ICs consuming a minimum of 110 watts of input power is obtained. For timing purposes and miscellaneous functions not specified, a total of 7,500 ICs is assumed.

The "D" type flip-flop ICs are housed on the dual row/column cards, which contain the microwave switching crosspoints, and are not considered here. The remaining ICs will be mounted on Cambion model #714-1115-01-00-00 9.75 inch x 9.25 inch general purpose IC boards with provisions for 140 input/output terminals. The 7 x 100 decoders require the largest number of integrated circuits after the "D" flip-flop. Each decoder requires 120 input/output lines which implies an individual card; thus 100 decoder cards are required. The remainder of the control logic can be more densely packaged and should be mounted adequately on 20 IC boards. Allowing for expansion and/or packaging difficulties, a total of 150 boards will be needed to house the control logic. Assuming a 1 inch thickness on the assembled board, the control logic volume required is approximately 8 cubic feet. A large volume of cabling can be

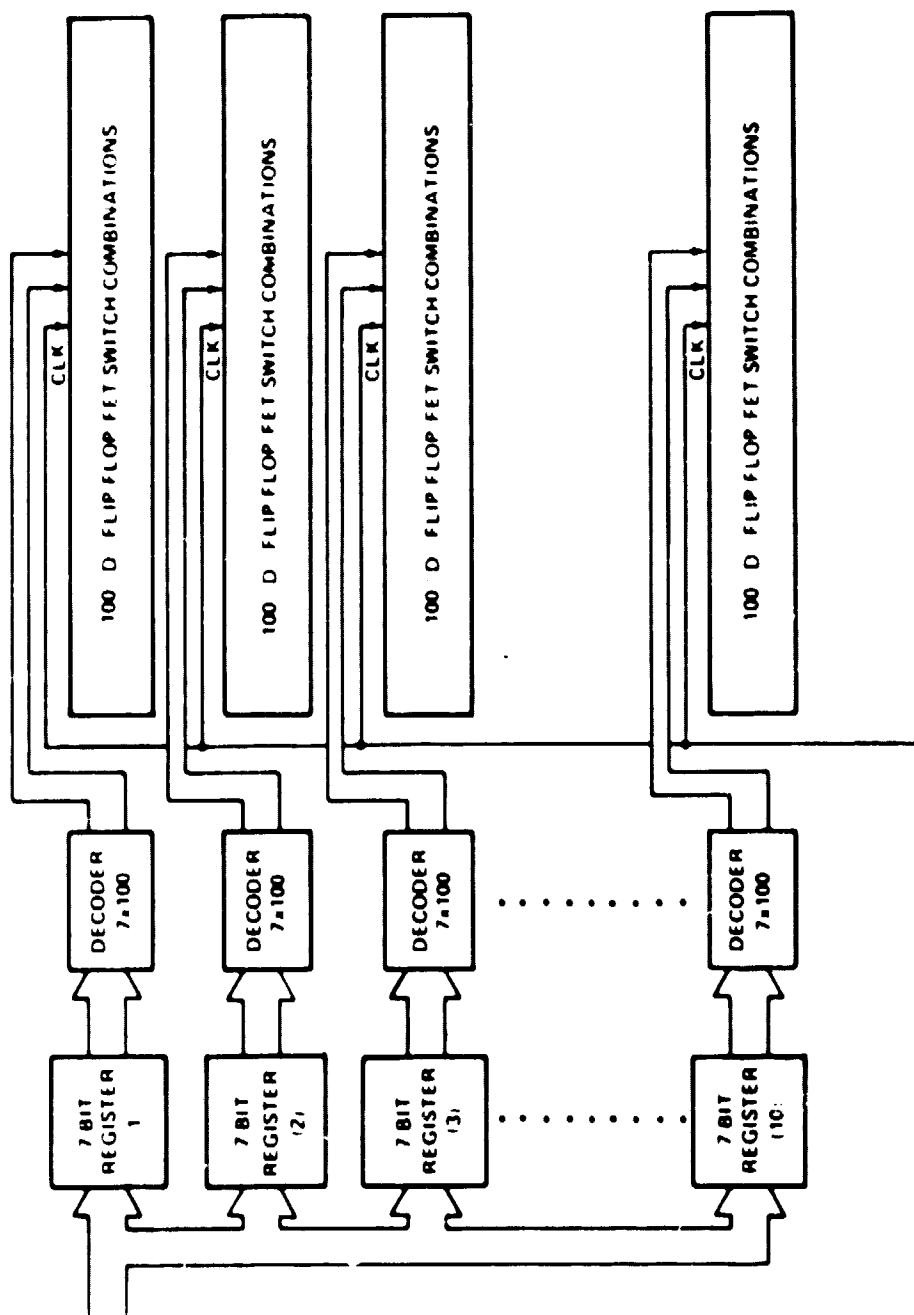


Figure 4-17. Level 3 Memory

Table 4-3
Sizing and Power Requirements

FUNCTION	PART NO.	PACKAGE	VOLTAGE (V)	TOTAL CURRENT/UNIT	CASE SIZE (in.)	QUANTITY REQUIRED	TOTAL INPUT POWER (W)
MEMORY	CDP1823	24 pin	10	$16 \mu\text{A}$	1.22×0.620	180	0.0288
BUS SEPARATOR	CDP1856	16 pin	10	$300 \mu\text{A}$	0.830×0.310	364	1.0000
HOLDING REGISTER	CDP1852	24 pin	10	$600 \mu\text{A}$	1.22×0.620	90	0.5400
DECODER	CDP1853	16 pin	10	$300 \mu\text{A}$	0.830×0.310	1500	4.5000
"D" TYPE FLIP/FLOP	SN54ALS74	14 pin	5	4 mA	0.785×0.310	5000	100.0000
SHIFT REGISTER	AM54LS299	20 pin	5	60 mA	1.05 x 0.385	11	1.3200
MICROPROCESSOR	CDP1804	40 pin	10	$100 \mu\text{A}$	2.09×0.600	1	0.0010
HEX INVERTER	MC14049S	16 pin	10	$10 \mu\text{A}$	0.785×0.350	17	0.0020
HEX BUFFER	MC14050B	16 pin	10	$10 \mu\text{A}$	0.758×0.350	220	0.0220

eliminated by housing the decoders in the same drawers as the microwave crosspoint cards.

4.5.2 Microwave Drawer

Ten drawers are required to contain the microwave switch. Each drawer contains 20 dual row/column microwave cards and ten 7 x 100 decoder cards. These cards plug into a multi-layer mother-board, which provides for the 1000 interconnections required between the D flip-flops and the 7 x 100 decoders contained in one drawer. This approach eliminates the necessity of bringing a 1000-wire cable to each drawer from the control logic. This concept is shown in figure 4-18.

4.5.3 Weight Estimate

The weights shown in table 4-4 have been compiled through catalog references, weight measurements, and in some cases, estimation. This table indicates that the 100 x 100 crossbar microwave switch will weigh about 1000 pounds.

4.6 CONCLUSIONS

A dual microwave row/column crosspoint has been designed. It is compatible with the best current microwave printed circuit card practice. Crosspoint size can probably be reduced. The control logic design is feasible using current technology. It represents catalog hardware rather than the latest in large scale integration and microcomputer technology. Significant reduction in parts count, weight, volume, and improvement in reliability could be realized by the use of custom LSI and IC boards for control logic.

As the design proceeded, it became apparent that a number of logic control circuits should be housed on the microwave boards or in the microwave drawers. This reduced card count, eliminated cables, improved timing, and saved weight and volume.

Considerable mechanical effort is required for the following:

- Card mechanical design
- Card mounting in drawers
- Spacing layout in drawers

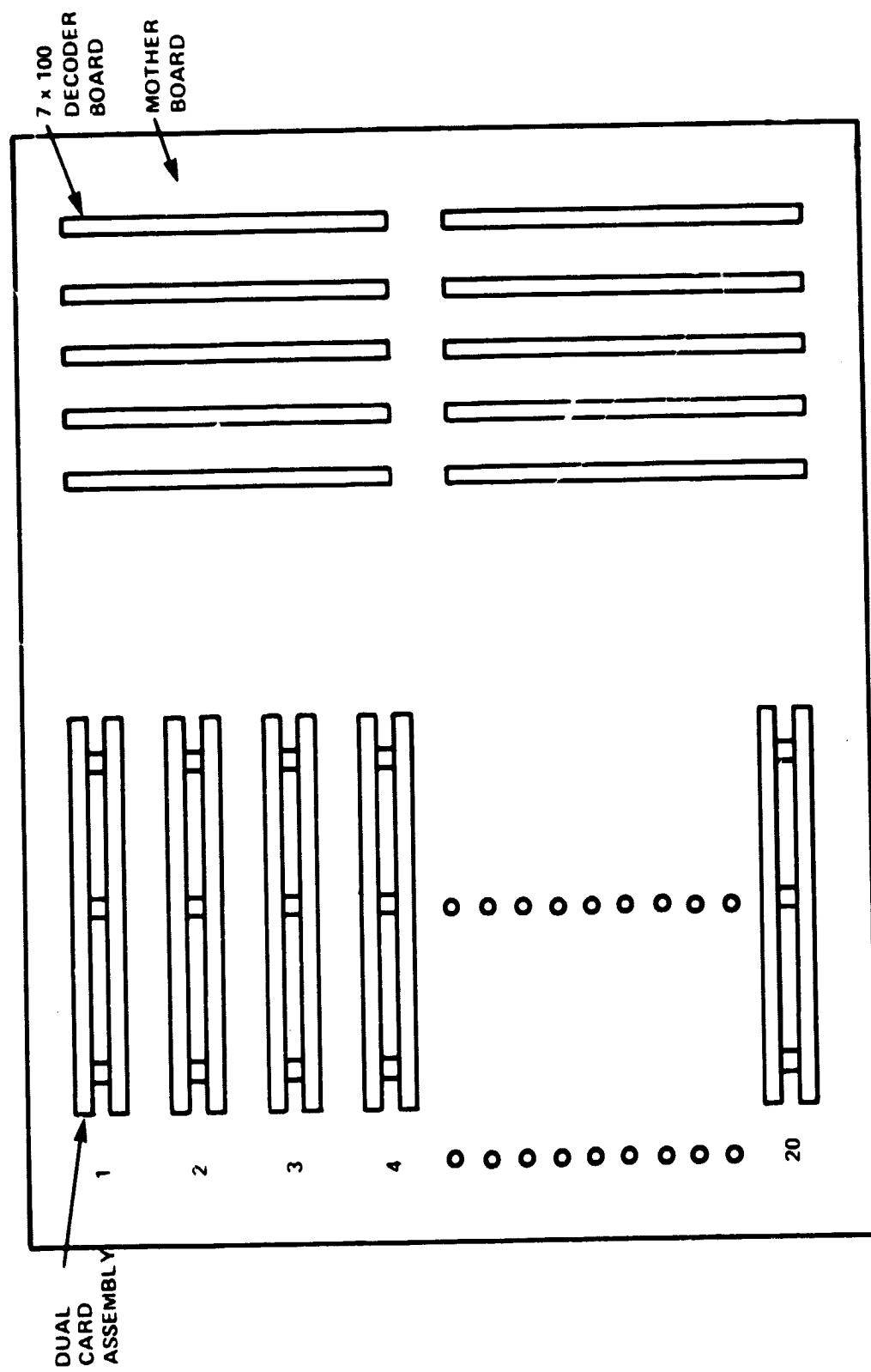


Figure 4-18. Decoders and Cross Point Cards in Drawer

Table 4-4
Weight Estimate

Quantity	Unit	Component	Total
7400	2 grams	Integrated circuits	32.6 lb
10,000	3 grams	Three-stage amplifier	66.1 lb
20,000	2 grams	Crosspoint terminations with mtg hardware	88.2 lb
200	0.627 lb	8½" x 11" dual card (Epsilon 10 PC board)	125.4 lb
140	0.793 lb	9.75" x 9.25" logic PC boards	111.0 lb
140	0.2 lb	PC board connectors	28.0 lb
10,000	3.5 grams	SMA cable & connector for dual card feed through connection	77.2 lb
10,000	4 grams	Mating connector for above	88.2 lb
6,000	3.5 grams	RF microstrip to coax connectors	46.3 lb
6,000	1.8 grams	Mating connector for above	23.8 lb
3,000	1 grams	Coax cable for above	6.6 lb
2,000	2.2 grams	Stand-offs for dual cards	9.7 lb
H/R	-	Locking material for conn.	5.0 lb
10	8 lb	Drawer (with mother-board) for microwave ass.	30.0 lb
1	10 lb	Interface drawer (remaining control logic)	10.0 lb
1	75 lb	Rack and mtg hardware for drawers	75.0 lb
A/R		Wiring harness	50.0 lb
		Miscellaneous	35.7 lb
		Total Weight	958.8 lb

- Cable and harness routing
- Installation procedures
- Drawer frame design
- Exterior frame design

All of these are beyond the scope of the current effort. The lack of such a program makes any weight estimate speculative and subject to very large errors.

REFERENCES

1. Joseph L. Katz, et al., "Final Report - Application of Advanced On-Board Processing Concepts to Future Satellite Communications Systems," MTR-3787, Vol. I, F19628-79-C-0001, Bedford, MA 01730: The MITRE Corporation, June 1979.
2. Raphael L. Edelman and Joseph L. Katz, "Bibliography, Application of Advanced On-Board Processing Concepts To Future Satellite Communications Systems," MTR-3787, Vol. II, F19628-79-C-0001, Bedford, MA, 01730: The MITRE Corporation, June 1979.
3. William T. Brandon and Brian E. White, "Executive Summary First Year Report: On-Board Processing Concepts for Future Satellite Communications Systems," MTR-3943, F19628-80-C-0001, Bedford, MA 01730: The MITRE Corporation, May 1980.
4. "30/20 GHz Mixed User Architecture Development Study - Final Report," NASA CR-159686, NAS 3-21933, Redondo Beach, CA 90278: TRW, Inc., Space Systems Division, October 1979, (I-7)-(I-11).
5. Gene L. Haviland and Al A. Tuszynski, "A CORDIC Arithmetic Processor Chip," IEEE J. Solid-State Circuits, Vol. SC-15, No. 1, (1980), 4-14.
6. P. Amadesi, et al., "On-Board Message Switching and Signal Processing," CSELT R. I. 78.07.302, ESTEC/No: 2872/76/NL/AK/SC, Torino, Italy: Centro Studi e Laboratori Telecommunicazioni S.p.A., July 1978.
7. John B. Anderson, et al., "Power-Bandwidth Performance of Smoothed Phase Modulation Codes," CRL-71, Hamilton, Ontario, Canada: Communications Research Laboratory, McMaster University, March 1980.
8. D. W. Davies and D. L. A. Barber, Communication Networks for Computers, New York: John Wiley & Sons, 1977.
9. Mischa Schwartz, Computer-Communication Network Design and Analysis, New Jersey: Prentice-Hall, Inc., 1977.
10. James W. Forgie, "Voice Conferencing In Packet Networks", ICC'80 Conference Record, Volume 2, (June 1980), pp. 21.3.1 - 21.3.4, New Jersey: IEEE Service Center, 1980.

11. Herbert J. Wolkstein and Jitendra Goel, "A 4-6 GHz Dualgate FET Amplifier," "Proceedings of the International Solid State Conference, February 16, 1978, p.26.
12. Reference Data for Radio Engineers, 6th edition, Indianapolis, IN: H. W. Sams & Co, 1975, p. 44-13.
13. M. Hoffman, "Satellite Microwave Switching Matrix Issues," WP22396, F19628-79-C-0001, Bedford, MA: The MITRE Corporation, 31 August 1979, pp. 114-115

APPENDIX A

TRAFFIC MATRICES

Table A-1 is a symmetrical traffic matrix based on a postulated customer premises terminal population over the eight regions of CONUS as arbitrarily defined in figure A-1. This appendix further explains the matrix derivation and begins to suggest how such a matrix might be used.

A.1 EXPLANATION OF NON-UNIFORM TRAFFIC MATRIX

Figure A-1 and table A-1 are configured to match the current MITRE baseline system for customer premises service (CPS). The baseline design contains eight fixed beams and eight scanning beams, each of which visits eight coverage areas during each frame. Although the traffic matrix is only 16 x 16, it represents a 72 x 72 matrix because each coverage area associated with a scanning beam is treated identically. That is, the first eight rows and the first eight columns are associated with fixed-beam areas, whereas the next eight rows and next eight columns are each associated with eight distinct areas covered by a scanning beam. Traffic is assumed to be uniform in each of the scanning beam coverage areas.

Other assumptions are: (1) traffic leaving a given coverage area is directly proportional to the population of customer premises terminals (CPTs) in that coverage area; (2) the number of CPTs in a given coverage area is proportional to the population in that coverage area; (3) the fraction of traffic from a given coverage area to any coverage area is proportional to the number of CPTs in the destination coverage area; (4) the total traffic leaving the largest population area, i.e., the New York City region, is normalized to unity, and the total traffic out of any other coverage area is reduced proportionately; (5) the traffic from coverage area A to coverage area B is assumed to be the same as the traffic from coverage area B to coverage area A; so with the natural ordering of rows and columns, the traffic matrix is symmetrical. These assumptions reflect and attempt to model CPS traffic flows which should be fundamentally different from normal telephone traffic where most of the calls are local.

The traffic matrix of table A-1 was derived from eight arbitrarily selected disjoint regions in CONUS and a CPT distribution postulated in TRW's first year's report for NASA/LeRC. [1]*

*This Appendix A reference is found on page 194.

Table A-1
Traffic Matrix Based on CONUS Customer Premises Terminal Population

		8 AREAS/REGION																	
		1 AREA/CITY								8 AREAS/REGION									
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	TOTAL	
NY	LA	CHIC.	DET.	PHILA.	SAN	WASH.	BOST.	REG.	REG.	REG.	REG.	REG.	REG.	REG.	REG.	REG.	REG.	6	
TMLS	700	505	505	321	306	228	220	210	1544	1326	923	862	849	560	310	262	9631		
TMLS AREA	700	505	505	321	306	228	220	210	193	166	115	108	106	70	39	33			
1 NY	.0727	.0524	.0524	.0333	.0318	.0237	.0228	.0218	.0200	.0172	.0120	.0112	.0110	.0073	.0040	.0034	1		
2 LA	.0378	.0378	.0240	.0229	.0171	.0165	.0157	.0145	.0124	.0086	.0081	.0079	.0079	.0052	.0029	.0025	.7214		
3 CHICAGO		.0378	.0240	.0229	.0171	.0165	.0157	.0145	.0124	.0086	.0081	.0079	.0079	.0052	.0029	.0025	.7214		
4 DETROIT			.0153	.0146	.0109	.0105	.0100	.0092	.0079	.0055	.0051	.0051	.0033	.0018	.0016	.4586			
5 PHILA.				.0139	.0103	.0100	.0095	.0088	.0075	.0052	.0049	.0048	.0032	.0018	.0015	.4371			
6 SAN FRAN.					.0077	.0074	.0071	.0065	.0056	.0039	.0036	.0036	.0024	.0013	.0011	.3257			
7 WASH. D.C.						.0072	.0069	.0063	.0054	.0038	.0035	.0035	.0023	.0013	.0011	.3143			
8 BOSTON							.0065	.0060	.0052	.0036	.0034	.0033	.0022	.0012	.0010	.3000			
9 REGION 1								.0055	.0048	.0033	.0031	.0030	.0020	.0011	.0009	.2757			
10 REGION 2									.0041	.0028	.0026	.0026	.0017	.0010	.0008	.2368			
11 REGION 5										.0020	.0018	.0018	.0012	.0007	.0006	.1648			
12 REGION 6											.0017	.0017	.0011	.0006	.0005	.1539			
13 REGION 8												.0017	.0011	.0006	.0005	.1516			
14 REGION 3													.0007	.0004	.0003	.1000			
15 REGION 7														.0002	.0002	.0554			
16 REGION 4																.0002	.0468		

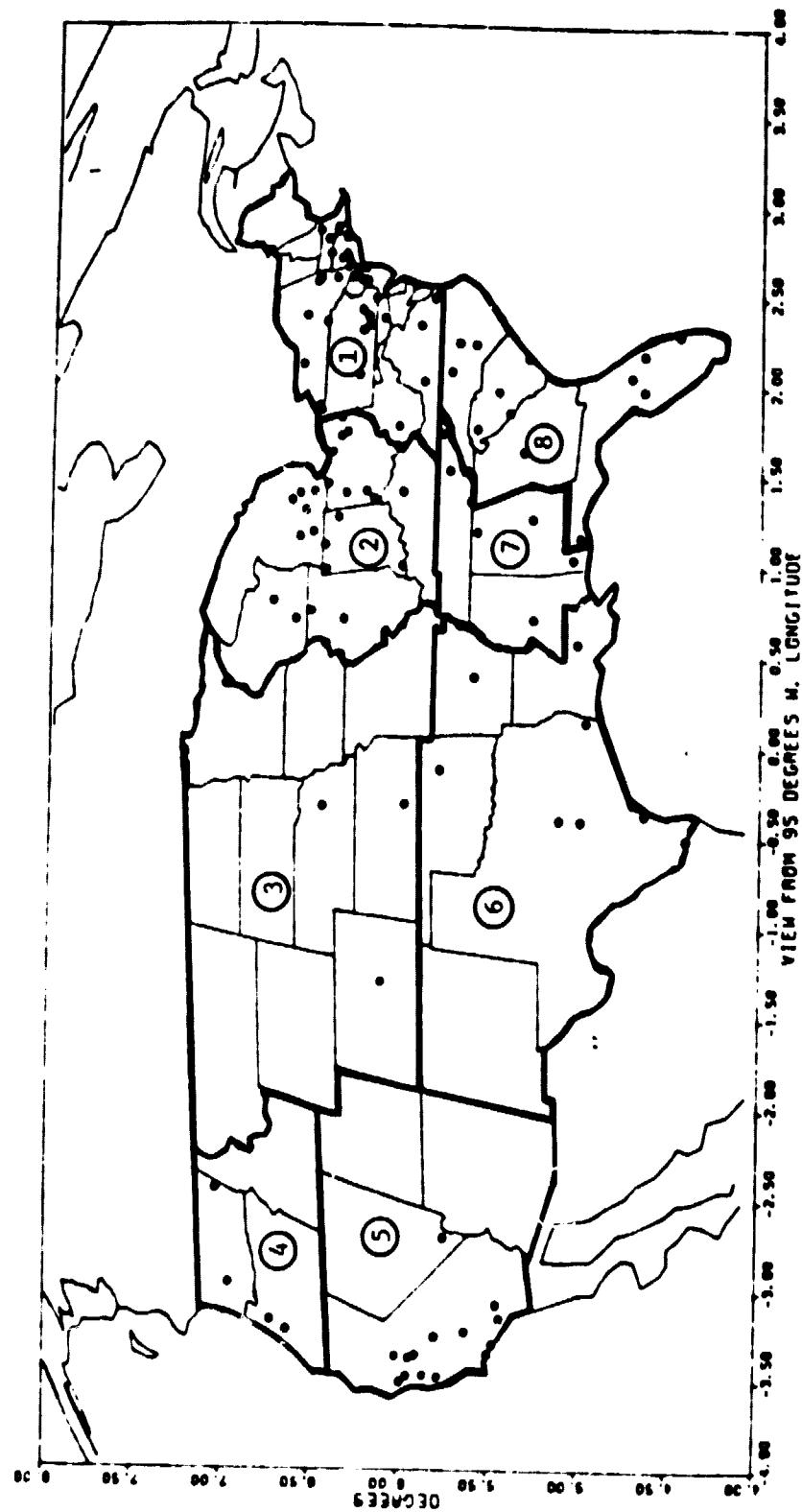


Figure A-1. CONUS Traffic Regions

For the normalization just defined, the elements of the first row (and the elements of the first column) sum to unity if the elements representing the scanning beams are each multiplied by eight for the eight scanning beam coverage areas. Note that the elements in any other row or column are scaled directly from row one or column one by the relative terminal population in the coverage area corresponding to that row or column.

The usefulness of this traffic matrix proposal is yet to be determined. However, some departure from the usual assumption of a uniform distribution of traffic should give satellite designers an idea of relative average traffic flows useful in implementing baseband switching concepts. For example, such a traffic matrix could indicate relative cumulative dwell times for different scanning beam areas.

A.2 APPLICATION OF TRAFFIC MATRIX TO BASEBAND PROCESSOR

For simplicity of illustration, consider a 3×3 version of the traffic matrix for the hypothetical areas A, B, and C which contain terminals in proportion to the numbers 13, 6, and 1, respectively. Listing the terminal populations by area according to the rows and columns and indicating that the traffic from a row i area to a column j area is proportional to the product of the number of terminals in the two areas, one obtains:

Area:	A	B	C	
Terminals/Area:	13	6	1	Row Sum
A: 13	169	78	13	260
B: 6	78	36	6	120
C: 1	13	6	1	20
Column Sum:	260	120	20	

The equivalent traffic matrix results from normalizing all the previous entries by the maximum row or column sum:

$$\begin{array}{c}
 \begin{array}{ccc}
 & \text{A} & \text{B} & \text{C} \\
 \text{A} & \left[\begin{array}{ccc}
 0.65 & 0.3 & 0.05 \\
 0.3 & 0.138 & 0.0231 \\
 0.05 & 0.0231 & 0.00385
 \end{array} \right] \\
 \text{B} & & & \\
 \text{C} & & &
 \end{array}
 \end{array}$$

This matrix, which results from an a priori traffic model based on terminal population, is interpreted as follows. Since area A has 13/20 of the total number of terminals, the aggregate collection of uplink or downlink antenna beams should cover area A 13/20 of the time, for example, on the average over many frame intervals. Within this aggregation of uplink dwell or slot times, 65%, 30%, and 5% of the uplink traffic from area A is destined for downlink areas A, B, and C, respectively. Similarly, within the downlink dwell or slot times, 65%, 30%, and 5% of the downlink traffic to area A originates from areas A, B, and C, respectively. These matrix elements indicate only the a priori traffic flows, e.g., $13 = 78/6 \approx 0.3/0.0231$ times as much traffic flows from area A to area B as from area B to area C, on the average.

Since user connections are continually being altered on a demand assignment basis, the average dwell or slot structure controlled by the baseband processor can be adapted to the actual traffic flows. Departures from the preprogrammed averages can be implemented automatically by responding to the actual traffic statistics measured on-board by the baseband processor.

For example, suppose the actual traffic matrix during a particular epoch was as follows:

$$\begin{array}{c}
 \begin{array}{ccc}
 & \text{A} & \text{B} & \text{C} \\
 \text{A} & \left[\begin{array}{ccc}
 0.596 & 0.323 & 0.0808 \\
 0.25 & 0.154 & 0.0577 \\
 0.0654 & 0.00769 & 0.00385
 \end{array} \right] \\
 \text{B} & & & \\
 \text{C} & & &
 \end{array}
 \end{array}$$

This matrix has been normalized by the maximum row or column sum, which in this case is the first row. Note that since the column sums are different from the corresponding row sums, the average fractional downlink dwell or slot times for a given area are different from those for the uplink. The control of the uplink and downlink beams could be adjusted according to these row and column sums.

REFERENCE

1. "30/20 GHz Mixed User Architecture Development Study - Final Report," NASA CR-159686, NAS 3-21933, Redondo Beach, CA 90278: TRW, Inc., Space Systems Division, October 1979, (I-7)-(I-11).

APPENDIX B

EFFICIENT ACCESS SCHEME FOR NON-REAL-TIME DATA

Low duty factor terminals with non-real-time data can utilize satellite channels very efficiently without any centralized control. This appendix addresses the feasibility of this concept by suggesting a set of subsystem parameters closely related to the baseline design and by sketching a conceptual implementation.

B.1 TERMINALS WITHIN FIXED BEAM COVERAGE AREA

For the moment, consider only terminals within a fixed beam coverage area connected by a single T1-rate channel. Let a super-slot consist of 2,048 frames of 1 ms duration, or 16,384 slots of 125 μ s duration. The 256 frames immediately preceding this super-slot will correspond to a super-subslot of 256 ms duration, roughly equal to a round-trip delay, which will be used to resolve contention among users of this channel.

Each user is assigned a unique integer from 1 to n. If a given user j has a packet to transmit, and it is his turn to do so, that user will transmit that packet if the current channel user i does not occupy the present super-subslot. Each packet is 2.048 s, or one super-slot, long. If user i occupies the present super-subslot, user j must not transmit until user i relinquishes the channel by not transmitting in a later super-subslot. Meanwhile, user i continues to transmit packets until his buffer is empty. At this time, user j will sense that the channel is not occupied and will transmit, unless his buffer is empty. In this event, the next user k, in some predetermined order of the n indices, will have the opportunity to transmit after sensing the channel idle one super-subslot later. This procedure is called the mini-slotted alternating priorities (MSAP) protocol.[1]*

Typical channel occupancy is indicated in figure B-1. The overhead in this multiple access scheme is primarily due to the fraction of the time super-subslots are not occupied. If every user always had exactly one packet to transmit when it was his turn, the overall channel efficiency would be $2.048 \text{ s} / (2.048 \text{ s} + 0.256 \text{ s}) = 8/9 \approx 89\%$ because there would be a continuous stream of alternating occupied super-slots and unoccupied super-subslots.

*References for Appendix B are listed on page 201.

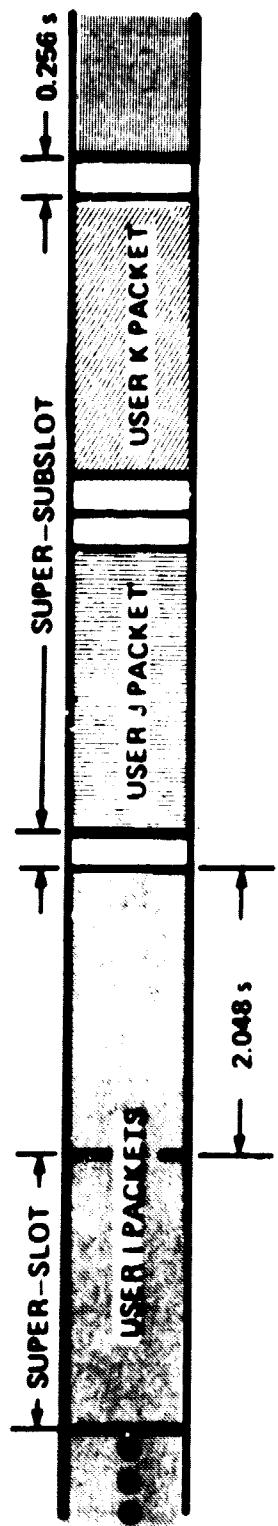


Figure B-1. Typical MSAP Satellite Channel Occupancy

A question that has not yet been addressed is how user j can determine whether a particular super-subslot is occupied. If user i and j are both in the same fixed beam coverage area, as assumed above, there really is no problem since all the channel traffic is broadcast on the downlink beam, and everyone in the net can identify unoccupied super-subslots, as well as traffic addressed to them.

Note that there is no need to regenerate this local network channel on-board the spacecraft other than to improve link performance by not repeating uplink interference. Of course, if the signal is regenerated, there will be a serial processing delay equivalent to the length of the buffer used to temporarily store demodulated data on-board. This would not affect channel utilization efficiency in any way, however.

If each user bursts at a rate of 1.544 Mb/s, an average data rate of approximately $R = 1.544/n$ Mb/s can be maintained at each of n terminals. The actual rate will be slightly less than R , since the unoccupied super-subslots induce overhead inefficiency. It should also be recognized that bursty terminal traffic is implied by a low terminal duty factor $1/n$. For example, twenty-four 64 kb/s terminals could be accommodated by one T1-rate satellite channel if each terminal had the capability of buffering several Mb of non-real-time data.

The average delay experienced by a single terminal in communicating a single 2.048 s packet of $(1.544 \text{ Mb/s})(2.048 \text{ s}) \approx 3.16 \text{ Mb}$ would be about $(n/2)(2.048 \text{ s}) \approx 24.6 \text{ s}$ near full channel loading, since the expected number of super-slots a terminal must wait before transmitting is $n/2$. This delay component dominates transmission and propagation delay and any on-board processing delay anticipated thus far, but arises from assuming that each terminal transmits a single packet in turn.

In practice, somewhat longer queueing delays are expected with heavy channel loading because in the MSAP protocol, a terminal transmits until its buffer is empty. On the other hand, with lighter traffic, considerably shorter average delays can be expected. The actual queueing delay formula in super-slots is given by: [2]

$$\begin{aligned}
 Q &= \frac{S}{2(1-S)} + \frac{a}{2} \left(1 - \frac{S}{n}\right) \left(1 + \frac{n}{1-S}\right) \\
 &= \frac{ng}{2(1-ng)} + \frac{a}{2} (1-g) \left(1 + \frac{n}{1-ng}\right), \quad g < \frac{1}{n}
 \end{aligned} \tag{B.1}$$

where $S = ng$, $g < 1/n$, is the expected channel throughput and g is the average terminal message load; $a = 0.256 \text{ s}/2.048 \text{ s} = 1/8$ corresponds to the one-way propagation delay between terminals in the net, i.e., a round-trip propagation delay through the satellite. The actual throughput is slightly less than S , when the overhead of the unoccupied super-subslots is taken into account. A graph of the queueing delay Q in packet slots of 2.048 s each is shown in figure B-2 for two values of n , the number of terminals in the single channel net. Observe that the queueing delay increases by less than a factor of two for forty-eight 32 kb/s users.

B.2 TERMINALS NOT WITHIN FIXED BEAM COVERAGE AREA

The next question that arises is whether a similar access scheme can be utilized for terminals within a scanning beam coverage area or for arbitrarily located terminals. Conceptually, the answer is yes, but the complexity of implementation and expected packet delay may increase for terminals in different coverage areas.

B.2.1 Terminals Within Scanning Beam Area

Since the data transmissions are to be accomplished in non-real time, there is no longer a requirement to have a short beam dwell interval and a fast beam hopping rate, e.g., 125 μs and 8 kh/s, respectively. A straightforward way of achieving roughly the same performance as in section B.1 would be to increase the terminal burst rate by a factor of eight and to subdivide the super-subslots into eight super-minislots each of duration 32 ms, as shown in figure B-3. The scanning beam would dwell on a given coverage area for a 32 ms interval every eight super-minislots, or every 256 ms. It would still take a terminal 2.048 s, or eight scanning beam periods to deliver essentially the same amount of data corresponding to one packet of section B.1, the difference being that this block of data is now delivered in eight packets of $8(1.54 \text{ Mb/s})(32 \text{ ms}) \approx 395 \text{ kb}$ each. The differences in channel bit overhead resulting from the need for burst synchronization is negligible because of the very long packet lengths compared to the tens of bits needed for synchronization.

The typical channel occupancy indicated in figure B-3 is analogous to that of figure B-1, except for the eight scanning beam coverage areas and the higher burst rate; channel efficiency, throughput, and delay performance (see equation (B.1) and figure B-2) would be nearly identical to that of section B.1.

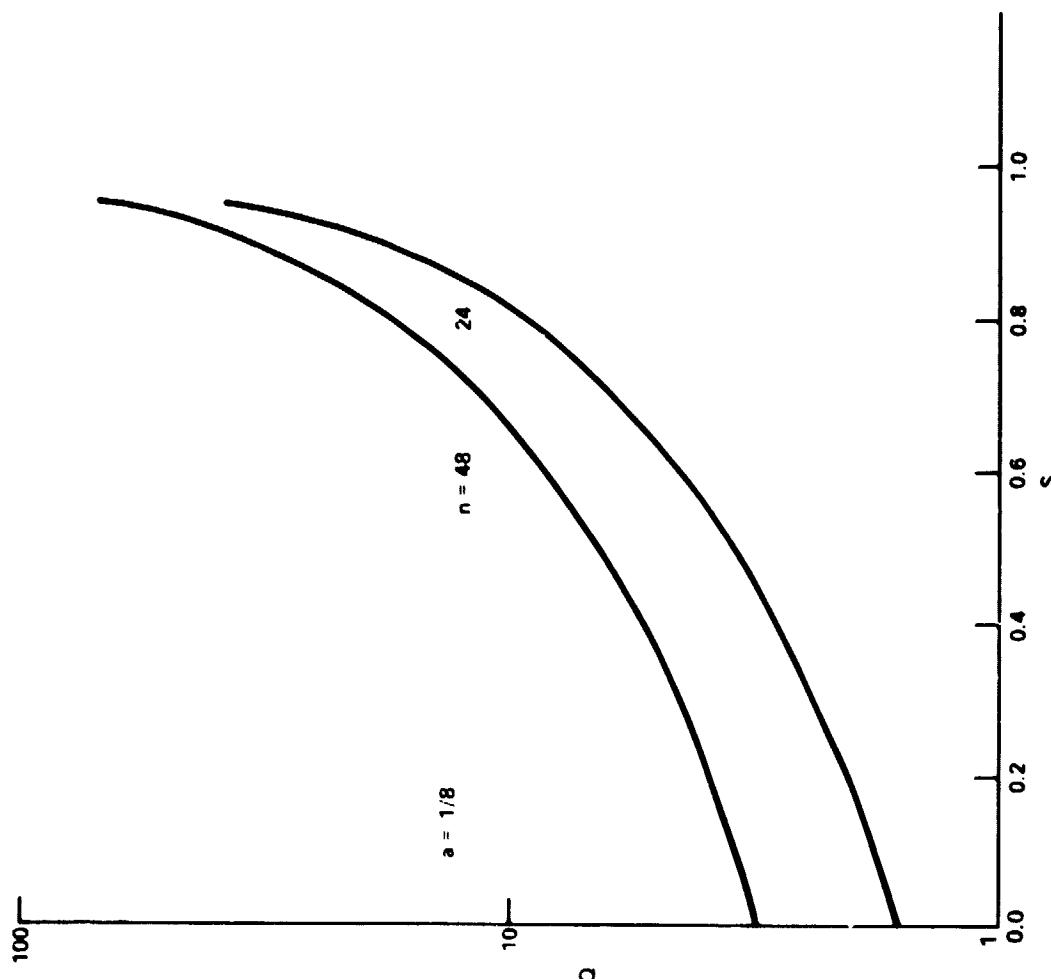


Figure B-2. Queuing Delay for MSAP Satellite Channel in Packet Slots of Duration 2.048 s

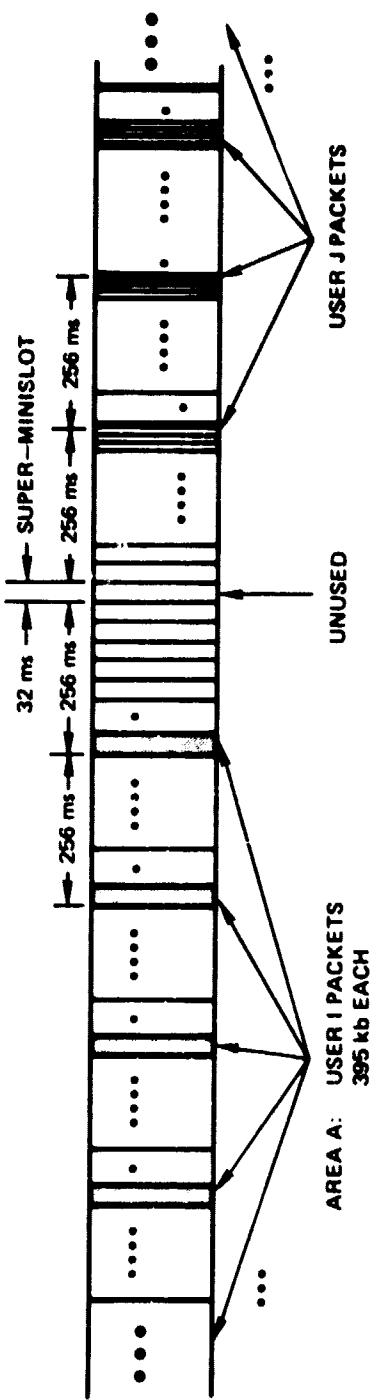


Figure B-3. MSAP Occupancy for Terminals in Scanning Beam Coverage Areas

B.2.2 Terminals in Different Coverage Areas

Now consider the most general case where user i in area A has a message for user j in area B. The need to recognize the destination address for satellite switching and to maintain the distributed control aspect of the access scheme forces on-board regeneration. It is convenient to think of the channel as being separated into an uplink channel for all source users like user i and a collection of downlink channels for all destination users like user j . The only information that must be broadcast to the uplink users on a downlink to area A to maintain the MSAP protocol is the occupancy status of the uplink channel, not the message packets themselves.

This information can be transmitted by having the satellite processor sense the unoccupied slots and by using a fixed beam, CONUS coverage, low rate, downlink orderwire channel which broadcasts the uplink access channel indices of only those channels which have just experienced an unoccupied MSAP contention slot. Uplink users of these channels would monitor this downlink broadcast and would have the opportunity to transmit packets on those channels immediately if it is their turn in the MSAP protocol. Packets successfully transmitted to the satellite would be routed to the appropriate downlink under the usual switching control procedures for non-real-time, or even real-time, traffic.

REFERENCES

1. Fouad A. Tobagi, "Multiaccess Protocols in Packet Communication Systems," IEEE Trans. Commun., Vol. COM-28, No. 4, (1980), 468-488.
2. Alan G. Konheim and Bernd Meister, "Waiting Lines and Times in a System with Polling," J. Assoc. Comp. Mach., Vol. 21, No. 3, (1974), 470-490.

APPENDIX C

ALL-DIGITAL DEMODULATION

Two basic approaches to all-digital demodulation are considered in this appendix. The first approach involves the straightforward A/D conversion of uplink signals followed by the multiplication and accumulation of b-bit quantized samples. The second approach utilizes CORDIC rotation which may have implementation advantages in realizing the multiplication functions.

C.1 A/D CONVERSION

The essential issues of analog-to-digital (A/D) conversion are the quantization level and sampling speed. Samples of the analog IF signal are "snapshots" of the real values at uniformly spaced discrete times. Each of these real values is approximated by one of a finite number of levels using a quantization rule. Each level is represented by a finite number of binary digits (bits).

C.1.1 Number of Quantization Bits Required

The received signal is assumed to be sampled, limited, and quantized according to standard techniques using the 2's complement representation of binary numbers with rounding. [1]* A central issue is the number of bits required to adequately represent a composite received waveform of bandwidth W containing a desired signal of bandwidth B, corresponding to the channel burst rate. Using the 1 b/s per Hz assumption of the baseline design, this implies that the number of FDMA user signals in the waveform is $N \approx W/B$. The usual Nyquist complex sampling rate W (complex samples/s) is assumed.

Let A be the received magnitude of the desired signal, let $N-1$ be the number of simultaneous in-band FDMA signals, each with received power I times larger than the desired signal, and let C be the clip level factor, i.e., the input waveform is clipped at a level C times above the measured rms level to maintain a certain fidelity criterion. If the received waveform has mean zero and variance σ_w^2 , if the variance of the channel noise is σ_z^2 , and if b is the number of quantization bits, not including sign, then the operative criterion is to select b large enough to ensure that the

*References for Appendix C are on page 211.

quantization noise variance $C^2 \sigma_w^2 / 12$ exp, $2b$ is much less than the variance of the desired signal alone plus channel noise, $A^2/2 + \sigma_z^2$. Defining the signal-to-noise ratio $A^2/2\sigma_z^2 = E_b B/N_o$, where E_b/N_o is the often used energy contrast ratio for the additive white Gaussian noise (AWGN) channel, and using the fact that

$$\sigma_w^2 = \frac{A^2}{2} (1 + (N-1)I + \sigma_z^2) \quad (C.1)$$

the above criterion can be rewritten as

$$b \geq \frac{1}{2} \log_2 \left[\frac{5}{6} C^2 \frac{\left(\frac{E_b/N_o}{W/B} [1 + (N-1)I] + 1 \right)}{\frac{E_b/N_o}{W/B} + 1} \right] \quad (C.2)$$

where a factor of 10 was used to indicate the "much less than" part.

For the sake of illustration, suppose that $C = 4$, $E_b/N_o = 10$, and that $W/B = N$. Table C-1 shows the minimum b values for $N = 1$, 8, and 64 (typical baseline design values) and $I = 1$ and 10. It should be emphasized that these results only represent the number of quantization bits required. A sign bit and any overflow bits needed during subsequent accumulation processing must be accounted for separately. However, the entries of table C-1 indicate the size of the multipliers required in the multiplication approach and the number of bits, excluding sign, in the input samples of the CORDIC rotation approach.

Table C-1

Minimum Number b of Quantization Bits Required, Excluding Sign and Overflow Bits, Under Various Baseline Design Scenarios Assuming $C = 4$, $E_b/N_0 = 10$ and $W/B = N$

<u>Number of User Signals N</u>	<u>Interference Factor I</u>	
	1	10
1	2	2
8	3	5
64	4	5

C.1.2 State-of-the-Art of A/D Converters

For purposes of this appendix, A/D converter data relative to the demodulation process, i.e., table C-1, has been extracted from section 3.9.4 of this report and developed here.

In the group demodulator approach, the sampling rate would be equal to the 100 MHz CPS bandwidth. According to table C-1, a 3 to 5-bit A/D converter is required since there are several simultaneous signals being demodulated within that bandwidth. A/D converters in this category are not beyond the state-of-the-art according to figure 3-28, section 3.9.4, e.g., TRW's 400 MHz, 5-bit A/D using silicon logic, but they may well consume too much power for satellite use. However, TRW has achieved low power (0.25 to 2 W), 4 to 8-bit, silicon (Si) A/D converters operating at 30 MHz and has predicted the feasibility of 4 to 8-bit A/D converters consuming only about 0.05 to 2 W at 1 GHz using gallium arsenide (GaAs) logic.

It is estimated that GaAs A/D converters of adequate quantization and speed could be developed for a 1990 satellite to permit all digital group demodulation. The power consumption of each A/D converter should be only a fraction of a watt. Assuming 0.3 W per A/D converter, based on a TRW prediction for 6 bits at 1 GHz, the thirty-two A/D converters in the all digital demodulation approach to the baseline design would consume a total of only 9.6 W. Recall that there are sixteen uplink antenna beams and the same 100 MHz IF bandwidth in each beam. Two A/D converters are required per beam, one for the in-phase group signal and one for the group signal in phase quadrature.

A more conservative scheme, which would bypass problems with A/D converters operating at 100 MHz sampling rates, is the individual mixing-down or frequency translation to baseband of each FDMA signal in the 100 MHz bandwidth. This typically would require sixteen 12.3 MHz A/D converters per uplink beam with only 2-bit quantizations. (See table C-1.) These parameters provide the other, and perhaps more realistic, benchmark leading to the total power consumption of the demodulators. In this case these A/D converters would consume 64 W total, even with a 4-bit quantization at 30 MHz.

In order to stay within the postulated weight and power budget mentioned in section 3, the demodulators should not require more than a few hundred watts. This would allow ample power for the signal acquisition and tracking and symbol synchronization circuitry as well as RF front-end hardware.

C.2 MULTIPLIER APPROACH

In general, the baseband modulations will be detected at the satellite receivers using two quadrature channels each with two multiplications per input sample, one corresponding to the local carrier reference mixing operation and one corresponding to multiplication by a baseband window. (See Appendix E.) Assuming a sampling rate of W complex samples/s, M uplink beams, and N channels per beam, the number of multiplications/s required is $4WMN$. If a single multiplication time is D , then the minimum number of multipliers required is

$$L = 4WMND. \quad (C.3)$$

If P is the power required to operate a single multiplier, then the total power for implementing the multiplication process is at least LP .

P is a function of b , D , and the logic family or other implementation employed in the multiplication. For a fixed b and for a given implementation, the faster the multiplication, the higher the power consumption. In a power-limited situation, twice the number of multipliers might be operated at half the speed with some mechanism for time-sharing the hardware to produce the same number of multiplications/s with less power. Similarly, in a speed-limited situation, several slower multipliers might be "pipelined" for achieving a required throughput multiplication rate.

The number of multipliers required and their attendant power consumption is estimated for the typical baseline parameters

$W = 100$ MHz, $N = 8$ or 64 , and $M = 16$. From equation (C.3), the minimum number of multipliers needed in any demodulation scenario is 5.12×10^{10} D and 4.096×10^{11} D for $N = 8$ and 64 , respectively. Note that if the multiplication time D is 10 to 100 ns, then hundreds or thousands of multipliers are required; the $N = 64$ channel/beam scenario may prove to be particularly troublesome.

It can be inferred from figure 3-26 of section 3.9.4 that GaAs multipliers offer attractive speed/power products for this application while Si devices cannot be considered serious contenders. In that figure, power data are presented for the number of multipliers per channel which are the least multiple of four not less than 4WD, cf. equation (C.3). This can result in a surplus of multipliers from the minimum required, but possible problems of multiplexing and/or switching among shared multipliers is avoided. However, pipelining and buffering of intermediate results is necessary.

The total power LP for implementing the multipliers is given in table C-2 for several baseline parameter combinations of interest using the GaAs data from figure 3-26 of section 3.9.4.

It is clear that 64 channels per beam is infeasible if the demodulators are limited to several hundred watts of power. As indicated in Table C-1, $b = 3$ is insufficient for $N = 64$. However, 8 channels per beam is quite possible even with 8 bits of quantization which is more than enough according to Table C-1.

Table C-2

Upper Bound on Total Power (W) Consumed by $b \times b$ GaAs Multipliers for $W = 100$ MHz and $M = 16$

		b		
		3	5	8
N	8	28.8	102.4	204.8
	64	230.4	819.2	1638.4

C.3 CORDIC ROTATOR APPROACH

Reduced to its essentials, in CORDIC rotation [2], each complex input sample from the A/D converter is to be multiplied by another complex number generated by the local mixing and windowing operations. Instead of doing this multiplication in the usual fashion, there may be advantages in recognizing that the local

reference signal has no variation in amplitude but only in phase, for constant envelope modulations (see Appendix E), and rotating the incoming sample in the complex plane by an angle θ specified by the local reference. Actually, θ is quantized and varies with each sampling instant. Each quantized rotation is accomplished by a series of quantized mini-rotations, where each successive mini-rotation represents a smaller angle. The sign of the mini-rotation is determined by the sign of the difference between the desired rotation angle θ and the current approximation to θ after the previous mini-rotation. The actual mini-rotation of the incoming complex sample is accomplished by successively changing the sign (or not) and interchanging the real and imaginary parts of shifted binary versions of the complex sample. The shifting is accomplished very simply by means of binary shift registers and is made possible by the fact that the mini-rotation angles possess tangents which are powers of one-half. A more precise description follows.

C.3.1 CORDIC Model

The basic function is to multiply the input sample (I, Q) by $G \exp(j\theta)$, where G is a small gain constant. The effect is a rotation by θ and slight stretching of the vector (I, Q) in the complex plane. It is easy to implement a rotation by $\hat{\theta} \approx \theta$ through a sequence of K mini-rotations involving angles of π , $\pi/2$, and angles which have tangents of $1/2$, $1/4$, etc., as listed in table C-3; $\hat{\theta}$ is given by

$$\hat{\theta} = r_{-1}\pi + r_0 \frac{\pi}{2} + \sum_{k=1}^K r_k \tan^{-1} 2^{-k} \quad (C.4)$$

where the coefficients, which are yet to be determined, assume the values $r_{-1}, r_0 = \{0, 1\}$ and $r_k = \{1, -1\}$, $k = 1, \dots, K$. The gain constant G_K approximating G can be expressed as (see table C-3)

$$G_K = \prod_{k=1}^K (1 + 2^{-2k})^{\frac{1}{2}}, \quad (C.5a)$$

and it can be shown that

$$G = \lim_{K \rightarrow \infty} G_K = 1.1646. \quad (C.5b)$$

This implies the need for only one overflow bit.

The traditional set of mini-rotation angles includes $\pm\pi/2$ and $\pm\pi/4$ instead of π and $\pi/2$ (or not). In 1975, C. M. Rader of the M.I.T. Lincoln Laboratory observed that the present set requires one less overflow bit while yielding equivalent quantization performance.

Table C-3
Approximate Mini-Rotation Angles and Gain Constants

k	$\tan^{-1} 2^{-k}$	K	$G_K \approx$
1	26.6°	1	1.118
2	14.0°	2	1.152
3	7.1°	3	1.161
4	3.6°	4	1.164

C.3.2 CORDIC Algorithm

The algorithm is initialized by converting θ to an angle θ_1 in the range $-\pi/4 < \theta_1 \leq \pi/4$, and (I, Q) to (I_1, Q_1) by changing (or not) the signs of I and Q . Given θ , the rules for choosing r_{-1} and r_0 are

$$r_{-1} = \begin{cases} 1, & 3\pi/4 < \theta \leq 7\pi/4 \\ 0, & \text{otherwise} \end{cases} \quad (C.6a)$$

$$r_0 = \begin{cases} 1, & \pi/4 < |\theta| \leq 3\pi/4 \\ 0, & \text{otherwise} \end{cases} ; \quad (C.6b)$$

this results in

$$\theta_1 = \theta - r_{-1}\pi - r_0 \frac{\pi}{2} . \quad (C.6c)$$

The values of (I_1, Q_1) are then determined according to table C-4; note that only sign changes and interchanges of real and imaginary parts are involved.

The remainder of the algorithm, steps 1 through K proceed according to the following rules. Given $-\pi/4 < \theta_1 \leq \pi/4$, select the K coefficients r_k and K-1 angles θ_{k+1} iteratively as

$$r_k = \begin{cases} +1, & \theta_k \geq 0 \\ -1, & \theta_k < 0 \end{cases}, \quad k = 1, \dots, K \quad (C.7a)$$

$$\theta_{k+1} = \theta_k - r_k \tan^{-1} 2^{-k}, \quad k=1, \dots, K-1. \quad (C.7b)$$

The mini-rotations are applied in an iterative fashion also by selecting the K modified complex samples (I_{k+1}, Q_{k+1}) by the rules

$$\left. \begin{aligned} I_{k+1} &= I_k - r_k Q_k 2^{-k} \\ Q_{k+1} &= Q_k + r_k I_k 2^{-k} \end{aligned} \right\}, \quad k = 1, \dots, K. \quad (C.8)$$

Note that a multiplication by 2^{-k} is accomplished by a shift of k places in a binary shift register. The r_k coefficients just indicate an addition or subtraction of components from the previous iteration.

Table C-4

Rules for Selecting (I_1, Q_1) Given r_{-1} , r_0 , and (I, Q)

r_{-1}	r_0	I_1	Q_1
0	0	I	Q
0	1	Q	-I
1	0	-I	-Q
1	1	-Q	I

C.3.3 Practical Considerations

In practice, of course, there is a rounding error associated with the implementation of equation (C.8). Based upon computer simulations of group demodulation, performed by L. N. Weiner and B. E. White at Lincoln Laboratory in 1975-76, it has been determined that a 6-bit quantization ($b = 6$), plus a sign bit, plus one overflow bit, and only four mini-rotations per rotation are required for an adequate representation. Thus, the real and imaginary parts can be adequately represented by 8 bits each, throughout the process of rotation and accumulation of results. It is possible that only 4 b and $K = 1$ or 2 would be required for demodulation of a single channel.

Since there are four complex rotations to be performed per input sample, corresponding to multiplying by the sine and cosine of the reference carrier plus multiplying by the sine and cosine of the reference window, the mini-rotation rate is $4KW$, where K is the number of mini-rotations per rotation and W is the complex sampling rate. If the rotations necessary for each input sample are to be accomplished with one set of shift registers, the worst-case shift-and-add time per iteration must be the reciprocal of this product. The mini-rotation time required for $W = 100$ MHz is 0.625 ns, 1.25 ns, and 2.5 ns, for $K = 4$, 2, and 1, respectively. Of course, two sets of registers might be used at half the rate, etc.

It remains to compute the power consumption for the CORDIC rotation approach as a function of the quantization size and the logic family and to compare these results with the expected power consumption for the multiplier approach. The following calculation is based on limited information and is therefore only an approximation.

Since the straightforward multiplication of two b -bit binary numbers requires $b-1$ additions, it is reasonable to expect a b -bit full adder to consume about the same power but $1/(b-1)$ -th the time of a b -bit multiplier. Thus, if an 8-bit GaAs multiplier performs in 6 ns with 0.4 W, as indicated in figure 3-26 of section 3.9.4, then an 8-bit full adder that is seven times faster at the same power level should be feasible. Unfortunately, $6 \text{ ns}/7 = 857 \text{ ps}$, so two such adders must be pipelined to meet the 625 ps mini-rotation time mentioned above. These adders might be able to operate slower, and consume less power, say, only 0.3 W per adder. Since there are two additions required per mini-rotation per channel, there are four adders consuming 1.2 W per channel. The GaAs multiplier is actually realized in a combinatorial fashion, and not straightforwardly, so this power estimate for the adder is tenuous.

The shifting of two 8-bit registers could be accomplished easily by Schottky diode field effect transistors (FETs). According to table 3-5 of section 3.9.3.1, this SDFL family can produce a 75 ps delay with 2.26 mW per gate. With a worst-case shift of 8 binary places consuming 590 ps, the 625 ps requirement could just be met. Since there are two registers required per mini-rotation, the shifting power per channel is only 36.16 mW.

This brings the power estimate for the CORDIC rotator approach to approximately 158 W for 128 channels. Although this is about 56 W more than the corresponding value for the multipliers, the power estimate for the CORDIC adders can easily be overestimated by 50%. Since this would bring the power consumption of the two approaches much closer together, a more detailed comparison seems warranted in future work.

REFERENCES

1. Alan V. Oppenheim and Ronald W. Schafer, Digital Signal Processing, Englewood Cliffs, NJ: Prentice-Hall, 1975, 408-417.
2. Gene L. Haviland and Al A. Tuszyński, "A CORDIC Arithmetic Processor Chip," IEEE J. Solid-State Circuits, Vol. SC-15, No. 1, (1980), 4-14.

APPENDIX D

SAWD AND CCD SIGNAL PROCESSING OVERVIEW

Since a thorough consideration of SAWDs and CCDs is beyond the scope of this report, an introductory overview is provided in this appendix. The interested reader is invited to explore the literature for further details. (Also, see section 3.9.3.1)

D.1 SAWD IMPLEMENTATION EXAMPLES

Many surface acoustic wave devices (SAWDs) have metallic interdigital transducers for launching or detecting an acoustic wave on the surface of a piezoelectric crystal. The physical shape of the transducer is used to control the response of the device, e.g., SAWD filters with fixed impulse or frequency response can be fabricated easily; some of the lowest cost SAWD filters can be found in commercial television receivers. Adaptive devices with programmable responses are possible but are more complex and more expensive, and some of these are still in the research stage. In contrast to the fixed SAWDs which use metallic or etched patterns, a family of programmable acoustoelectric devices that are insensitive to temperature variations exists. [1]*

Some of the general characteristics of SAWDs are listed in table 3-2 of section 3. The potential for low volume, weight, and cost is particularly attractive. This is explored here briefly for several SAWD applications.

D.1.1 Generation of MSK Modulation

Minimum Shift Keying (MSK) is one of the bandwidth and power efficient modulations highlighted in MITRE's first year report for NASA/LeRC. [2] Although MSK is an increasingly popular modulation, it is being surpassed in performance by the advanced modulation/coding schemes of Appendix E. On the other hand, the implementation of MSK is well understood, and much remains to be done in achieving practical realizations of some of the modern techniques. In particular, the use of SAWDs in the generation and detection of various modulations is of interest.

*References for Appendix D are listed on page 219.

The dramatic simplification possible in the generation of MSK with the introduction of SAWDs is evident in figure D-1.

Conventional schemes are rather ponderous, though effective. The functionally equivalent SAWD scheme employs a bi-phase modulated pulsed sinusoid at one signaling frequency to drive a SAWD filter with an impulse response consisting of a pulsed sinusoid at the other signaling frequency. Details of this continuous phase shift modulated (CPSTM) waveform implementation can be found in the literature.[3]

D.1.2 UHF Oscillators

Since SAWDs operate in the VHF/UHF range, considerable savings in weight and volume may be achieved by RF (or even IF) operation at several hundred MHz.* SAWDs employed as oscillators can achieve a 20:1 advantage over more conventional approaches, as suggested by figure D-2.

D.1.3 Reflective Array Compressors

The reflective array compressor (RAC), invented by R. C. Williamson of the M.I.T. Lincoln Laboratory, is one of the earlier signal processing devices using SAWs. This device is particularly useful for the generation or matched filter detection of a chirp signal, i.e., an elementary waveform which has an instantaneous frequency that varies linearly with time. Chirp waveforms can be used as a means of spreading the signal spectrum for achieving a multiple access interference rejection capability, for example.

The RAC shown in figure D-3 can be used to generate a "down" chirp by applying a time-domain impulse at the input transducer, or as a matched filter for an "up" chirp. In either case, the incoming higher frequencies propagate shorter distances into the device until their wavelengths match the spacings of the etched grooves or "fingers". Lateral reflections across the device occur at these points and the waves propagate in the opposite direction toward the

*Magnetostatic-wave (MSW) devices that perform some of the functions of SAWDs are becoming available. MSW devices operate directly at microwave frequencies of 1 to 20 GHz with bandwidths of 0.5 to 1 GHz. However, delays of only 10 to 500 ns are possible. (J. D. Adam, M. R. Daniel, and D. K. Schroeder, "Magnetostatic-Wave Devices Move Microwave Design Into Gigahertz Realm," Electronics, Vol. 53, 8 May 1980, 123-128.)

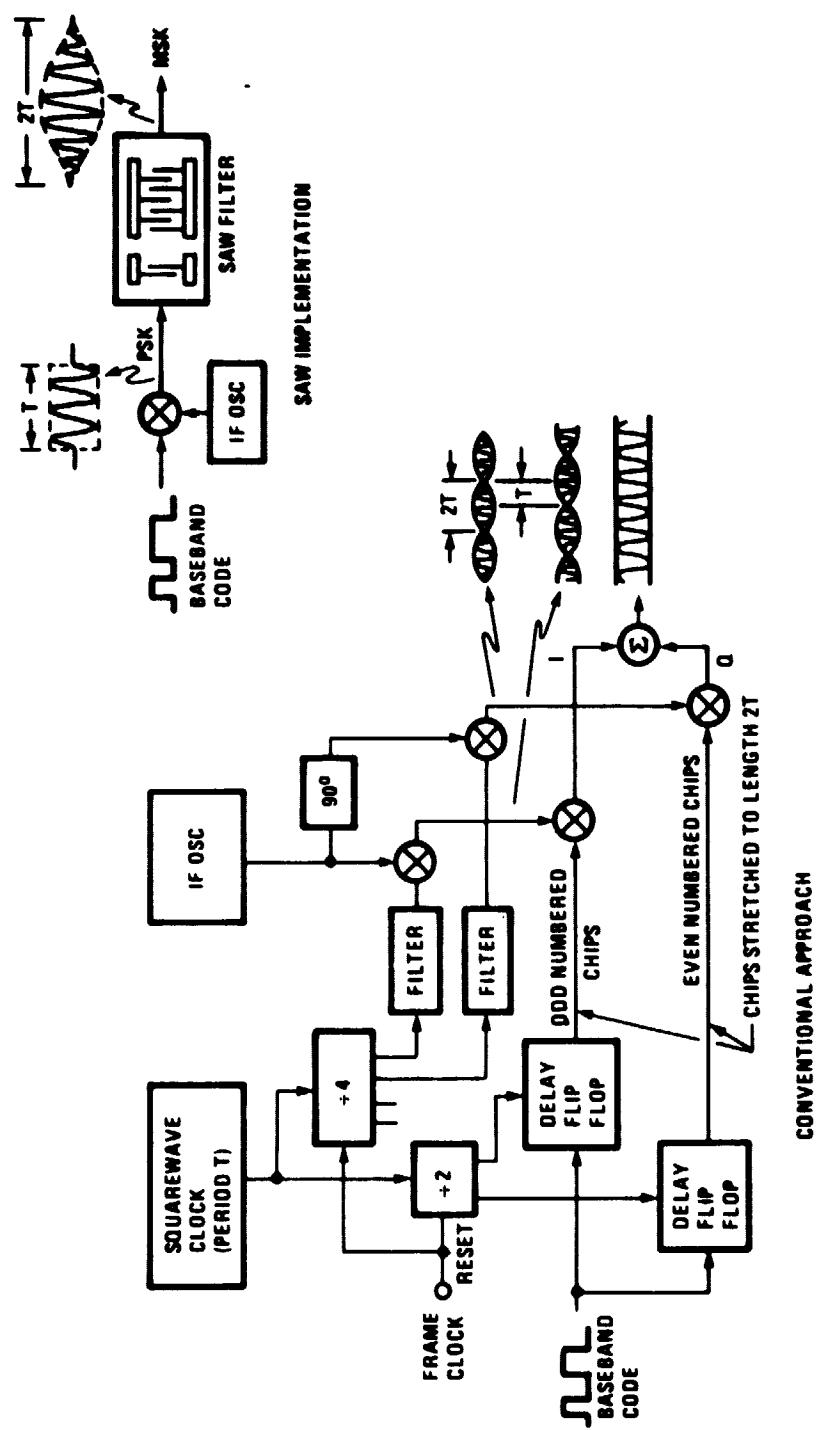


Figure D-1. Generation of MSK

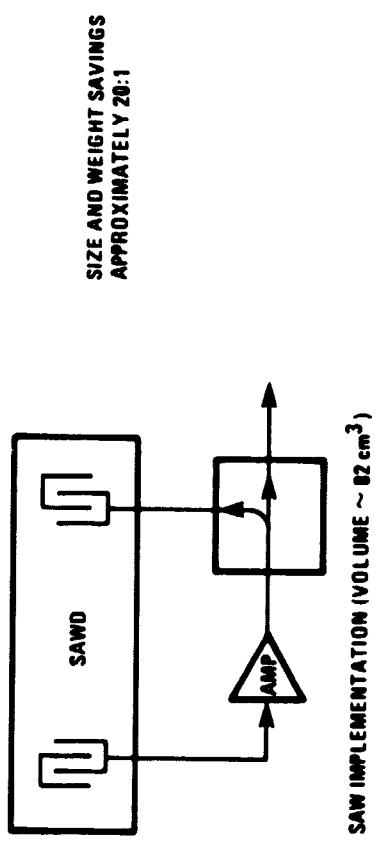
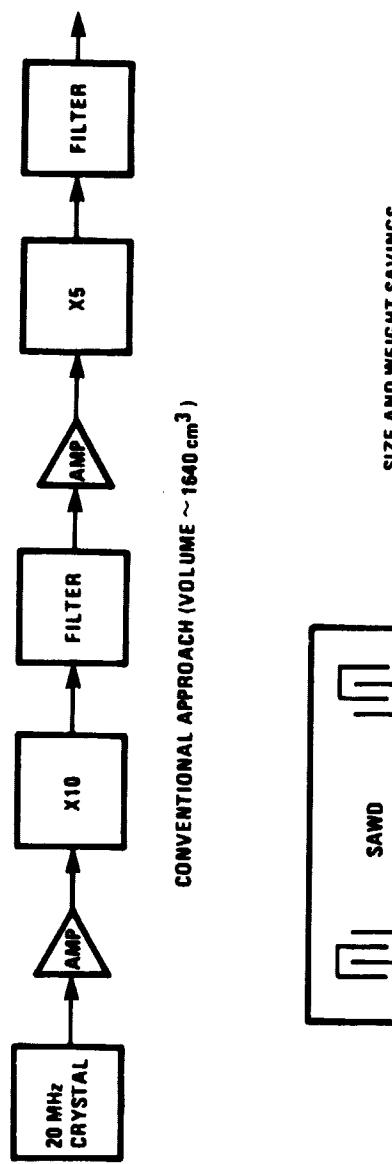


Figure D-2. Typical 1 GHz Oscillators

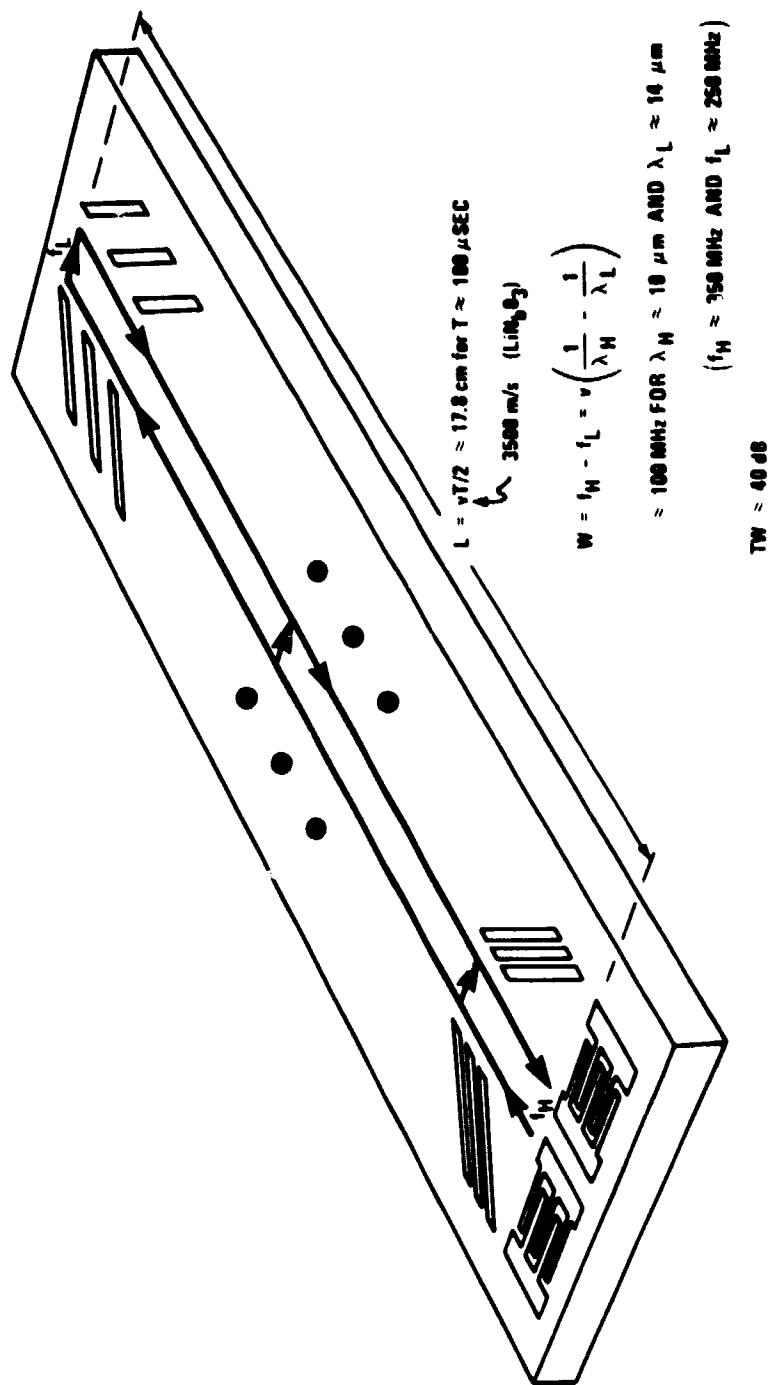


Figure D-3. Reflective Array Compressor (RAC) Example

output transducer. Longer delays are incurred by the lower frequencies. Thus, a frequency that decreases linearly with time results from an impulse, and an impulse delayed by twice the RAC length divided by the SAW propagation speed v results from an up chirp.

Maximum time dispersions (delays) of $T = 100 \mu s$ or more are attainable with RACs of 18 cm or more in length for a common piezoelectric material, lithium niobate, which has a propagation speed of approximately 3500 m/s. The state-of-the-art in growing sufficiently pure crystals limits the maximum delay. Wavelengths are limited on the low side by the lithographic state-of-the-art of several microns (μm) in etching fingers. Hence, maximum frequencies of $f_H = v/\lambda_H$ are limited to hundreds of MHz. The RAC bandwidth is the difference between the highest and lowest tunable frequencies, or $W = f_H - f_L = 100$ MHz, in this instance. The time-bandwidth product, or so-called processing gain, of the RAC shown is $TW = 10^4 = 40$ dB.

RACs can also be used as components for realizing chirp-z transforms for frequency analysis [4], and for fast, coherent synthesizers for frequency hopping waveforms.[5]

D.2 CCD SIGNAL PROCESSING

Charged coupled devices (CCDs) are used in a wide variety of signal processing applications in communications technology, notably in the single IC-chip filtering of analog, sampled-data signals.[6] The basic principle consists of the serial, weighted transfer of analog quantities of electrical charge at discrete times corresponding to uniformly spaced sampling instants of the continuous input signal. As with SAWDs, the usual process of A/D conversion is avoided.

An example of higher-order signal processing state-of-the-art using only CCDs was produced recently for the U.S. Army by Texas Instruments, Inc.[7] This development resulted in a general purpose IC chip of CCDs capable of performing a 512-point discrete Fourier transform (DFT) using the chirp-z transform algorithm. The chip includes four CCD recirculating transversal filters for performing convolutions, four 10-bit multiplying digital-to-analog (D/A) converters (MDACs) for performing multiplications, CCD clock logic and drivers, four amplifiers, and control logic, buffers, and other support hardware. The processor operated at a clocking rate of up to 1 MHz with more than 60 dB of dynamic range; power dissipation at 1 MHz was 475 mW. The complete IC chip measures 6.1 mm \times 5.5 mm.

D.3 HYBRID SAWD/CCD ACCUMULATING CORRELATOR

When using a SAWD as a matched filter for elementary binary chirp modulation, for example, it can be shown that the ideal, coherently detected, matched filter output in the absence of noise is a continuous time function of the complex form

$$f_{\omega}(t) = e^{j(\omega + \frac{mT}{2})(t-T)} \cdot \frac{\sin \left[\frac{m(t-T)(t-1/2T)}{2} \right]}{\frac{m(t-T)T}{2}}, \quad (D.1)$$

$0 \leq t \leq 2T,$

where ω is the radial carrier frequency, m is the radial frequency slope of the (up) chirp, and T is the chirp duration. The received signal is demodulated by sampling the filter output at $t = T$ (note that $f_{\omega}(T) = 1$) and sensing the sign or magnitude of the result, depending on whether information is coded into the phase or the slope of the elementary chirps.

In any event, since the $f_{\omega}(t)$ is generally a wideband signal which may be difficult to sample with sufficient precision for coherent detection, a form of non-coherent (envelope) detection is often employed in conjunction with two SAWDs, one matched to an up chirp and one matched to a down chirp. This reduces the problem to the detection of $|f_{\omega}(t)|$, which eliminates the ω dependence, as seen from equation (D.1). Although this may simplify the receiver by relaxing the timing requirements, the performance is degraded by up to several decibels in E_b/N_0 compared with the coherent detection of binary antipodal signals in AWGN. The actual degradation depends on the cross-correlation coefficient of the elementary signals, which in turn depends on the time-bandwidth product of the chirp. For time-bandwidth products of ten or more, the up and down chirps are nearly orthogonal, i.e., their cross-correlation coefficient can be taken as zero, for a 3-dB loss compared to antipodal signals.

In general, for these reasons it may be desirable to use a hybrid SAWD/CCD matched-filter detector or correlator to demodulate communications signals. The SAWD provides the wideband input capability and the CCD slows down the output for easier detection. A single device like this is under development at the M.I.T. Lincoln Laboratory.[8]

REFERENCES

1. John H. Cafarella, "Acoustoelectric Surface Wave Devices for Programmable Signal Processing," *Proc. Intern. Symp. Circuits Systems*, IEEE, 1978, 392-399.
2. Joseph L. Katz, et al., "Final Report - Application of Advanced On-Board Processing Concepts to Future Satellite Communications Systems," MTR-3787, F19628-79-C-0001, Vol. I, Bedford, MA 01730: The MITRE Corporation, June 1979.
3. W. Richard Smith, "SAW Filters for CPSM Spread Spectrum Communication," *Proc. Ultrasonics Symp.*, Phoenix, AZ: IEEE, 26-28 October 1977, 524-528.
4. Mervyn A. Jack, Peter M. Grant, and Jeffrey H. Collins, "The Theory, Design, and Applications of Surface Acoustic Wave Fourier-Transform Processors," *Proc. IEEE*, Vol. 68, No. 4, (1980), 450-463.
5. J. M. Hannah, P. M. Grant, and J. H. Collins, "Fast Coherent Frequency Hopped Waveform Synthesis Using SAW Devices," *Proc. Ultrasonics Symp.*, Annapolis, MD: IEEE, 29 September - 1 October 1976, 428-431.
6. Allen Gersho, "Charge-Transfer Filtering," *Proc. IEEE*, Vol. 67, No. 2, (1979), 196-218.
7. Robert C. Pettengill, Pat W. Bosshart, and Michiel de Wit, "CCD Circuits for Signal Processing," DELET-TR-76-1386-F, DAAB07-76-C-1386, Dallas, TX 75222: Texas Instruments, Inc., April 1979.
8. Richard W. Ralston, Daniel L. Smythe, and Ernest Stern, "A SAW Accumulating Correlator with CCD Readout," *Proc. Ultrasonics Symp.*, IEEE, 1979, 771-775.

APPENDIX E

BANDWIDTH AND POWER EFFICIENT MODULATIONS

Considerable discussion of bandwidth and power efficient modulations appears in MITRE's first year reports to NASA/LeRC. [1,2]* This appendix updates and expands the previous material.

E.1 CODED, CONSTANT ENVELOPE MODULATIONS

Attention is still focused only on constant envelope (CE) modulations. It is true that there are good non-CE modulations for use on bandlimited and nonlinear channels, and that CE modulations lose the CE property on such channels. [3] Nevertheless, the trend toward smaller terminals and the premium on transmitter power amplifier efficiency for such terminals favors CE modulations. This prevents distortions in the transmitted signal by the highly nonlinear power amplifier operating in the saturated Class C mode for high efficiency. Furthermore, it has been shown that coded CE waveforms with very attractive bandwidth and power efficiencies exist. [4] It remains to discover some of the better codes and to explore further the synchronization requirements of these advanced modulations.

The improved performance promised is extracted with increased demodulation complexity. This system trade-off is expected to be favorable to the extent that receiver complexity can be handled with low cost VLSI circuitry. Practical implementation difficulties may continue to provide opportunities for other waveform designs, however, such as simpler CE modulations and non-CE modulations. For instance, in some of the advanced CE waveforms discussed, the relatively small phase changes inherent in the modulation tends to make synchronization more difficult. Added receiver hardware required could be traded for linear power amplifiers at the transmitters, for example.

A general principle in communication theory is to unify the design of the source/sink, coder/decoder (CODEC), and modulator/demodulator (MODEM) so as to be more effective in connecting end users with equipment and techniques well-matched to the problem and the channel. In particular, one should not design the CODEC independently of the MODEM. This was the usual practice up to a few years ago until it became better known that several decibels of increased coding gain could be obtained by making soft

*References for Appendix E are listed on page 234.

decisions rather than hard decisions at the demodulator. Currently, a small revolution in modulation and coding is leading to coding gains without bandwidth expansion. With higher-level (non-binary) symbol alphabets and elementary baseband waveforms that possess smoothness properties of at least one level of frequency continuity and that extend over more than one underlying symbol interval, coding gains are possible in smaller bandwidths, i.e., bandwidth and power efficiency can be improved simultaneously. As is described below, researchers in Canada (J. B. Anderson of McMaster University) and Sweden (C.-E. W. Sundberg and T. Aulin of Lund University and N. Rydbeck of SRA Communications) are treading dangerously close to the Shannon limit!

E.2 THEORETICAL DEVELOPMENT

Consider a constant envelope (CE) signal where there is the opportunity for an underlying change in phase at the beginning of every elementary signaling or symbol interval of fixed duration T . If E and ω denote the signal energy in this interval and the radial frequency of the RF carrier, respectively, then the received signal is of the form

$$s(t) = \sqrt{\frac{2E}{T}} \cos(\omega t + \phi(t)). \quad (E.1)$$

In the subclass of CE signals considered here, the signal phase carrying the information can be expressed as

$$\begin{aligned} \phi(t) = \pi h \left(\sum_{i=0}^{n-L-1} a_i + 2 \sum_{i=n-L}^{n-1} a_i \int_{it}^t f(t-iT) dt \right), \\ (n-1)T < t \leq nT, \quad n = 1, 2, \dots \end{aligned} \quad (E.2)$$

where h is a constant parameter often called the modulation index,

a_i denotes the i th data symbol selected from the alphabet $\{\pm 1, \pm 3, \dots, \pm (M-1)\}$ for alphabet size M even, or $\{0, \pm 2, \pm 4, \dots, \pm (M-1)\}$ for M odd, and $f(t)$ is a baseband frequency pulse which is zero outside the interval $(0, LT)$, given a positive integer pulse length L , measured in elementary symbol intervals, and normalized by

$$\int_0^{LT} f(t) dt = \frac{1}{2}. \quad (E.3)$$

As is shown later, bandwidth efficiency, i.e., compact baseband signal spectra with rapid sidelobe roll-off, can be achieved by a combination of small h , large L , and a smooth baseband frequency

pulse shape, preferably with at least one continuous derivative. Since such bandwidth efficiency may be achieved at the expense of detection or power efficiency, one is concerned with the E_b/N_0 required to attain a given probability of bit error, i.e., error rate, with these waveforms in the presence of additive white Gaussian noise (AWGN). Alternatively, one can attempt to find a single parameter F_0 which characterizes the modulation by specifying the maximum number of bits per channel use that can be transmitted for a given E/N_0 with arbitrarily small probability of error by means of some random code. The latter more modern approach, which is described in reference [1], is adopted here in order to present some results of Anderson and Sundberg, et al.[4]

The power efficiency of these CE waveforms can be estimated by considering the probability of selecting the wrong coded sequence of underlying phase changes at a maximum likelihood receiver. Legitimate codewords comprise some strict subset of all M^n possible sequences of length n from the alphabet of size M . Suppose the i th data sequence codeword of length n

$$A_i(n) = a_{0i} a_{1i} \dots a_{n-1,i}$$

was sent. This codeword would define a corresponding transmitted signal waveform $s_i(t)$ according to equations (E.1) and (E.2). The received waveform, corrupted by AWGN, would be mapped into the nearest possible signal $s_j(t)$ in the signal space which corresponds to the codeword $A_j(n)$. Assuming that the code is selected at random and that the codewords are chosen independently with the same specified probability distribution, the error probability is union bounded by

$$P_{\text{em}} \leq \overline{m} P_{\text{e2}} = \overline{m} \bar{Q} (d_{ij} / \sqrt{2N_0}), \quad i \neq j \quad (\text{E.4a})$$

where the number of codewords in the code is

$$\overline{m} < M^n \quad (\text{E.4b})$$

where the overbar denotes a statistical expectation over all possible codes and all possible distinct codeword pairs (A_i, A_j) , where d_{ij} is the Euclidean distance between $s_i(t)$ and $s_j(t)$ in the signal space, and where

$$\begin{aligned} Q(x) &= \frac{1}{2} \operatorname{erfc} (x/\sqrt{2}) \\ &= \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(-z^2/2) dz. \end{aligned} \quad (\text{E.4c})$$

The distance d_{ij} is obtained from

$$\begin{aligned}
 d_{ij}^2 &= \int_0^t (s_i(t) - s_j(t))^2 dt \\
 &= \int_0^t s_i^2(t) dt + \int_0^t s_j^2(t) dt - 2 \int_0^t s_i(t)s_j(t) dt \\
 &\approx \frac{2E}{T} \left(t - \int_0^t \cos[\phi_i(t) - \phi_j(t)] dt \right), \\
 (n-1)T < t &\leq nT, \quad n = 1, 2, \dots
 \end{aligned} \tag{E.5}$$

from equations (E.1) and (E.2), where the approximation holds for large carrier radial frequency ω . Equation (E.5) shows how the Euclidean distance in the signal space depends on the phase difference $\Delta\phi_{ij}(t) = \phi_i(t) - \phi_j(t)$ between the two coded waveforms. The precise value of $\int_0^t \cos \Delta\phi_{ij}(t) dt$ requires detailed calculation, in general.

Let the code rate be defined as

$$R = \frac{\log_2 m}{n} \text{ (bits/symbol)} \tag{E.6a}$$

i.e., for m equally likely codewords, $\log_2 m$ bits of information are transmitted in n symbol intervals. Therefore, in the absence of noise, an average of R information bits per symbol is received. Now, following Anderson, let a bound parameter R_0 be defined as

$$R_0 = -\frac{\log_2 P_{e,i}}{n} \text{ (bits/symbol).} \tag{E.6b}$$

Then, substituting equation (E.6) into equation (E.4a) yields

$$p_{\text{ea}} \leq 2^{-n(R_o - R)} \quad (\text{E.7})$$

which states that there exists a code of rate R yielding an arbitrarily small message error probability, exponentially decreasing with increasing sequence length n , provided $R < R_o$.

E.3 PERFORMANCE EXAMPLES

Anderson has computed R_o for two representative families of bandwidth and power efficient modulation in the CE subclass under consideration. These two families, whose frequency pulses he denotes by LD and LRC, are defined as follows.

o LD Frequency Pulses

An LD frequency pulse has $L-2$ continuous derivatives, extends over the L symbol interval $(0, LT)$, and has an integral of $1/2$; cf. equation (E.2). The analytical expressions for the first four pulses in the family, sketched in figure E-1, are

$$\text{OD: } f(t) = \frac{1}{2} u_0(t) \quad (\text{impulse of area } 1/2 \text{ located at the origin})$$

$$\text{1D: } f(t) = \frac{1}{2T} [u_{-1}(t) - u_{-1}(t-T)] \quad (\text{rectangle of height } 1/2T \text{ and width } T)$$

$$\text{2D: } f(t) = \frac{1}{2T^2} [u_{-2}(t) - 2u_{-2}(t-T) + u_{-2}(t-2T)] \\ (\text{isosceles triangle of height } 1/2T \text{ and base } 2T)$$

$$\text{3D: } f(t) = \frac{1}{2T^3} [u_{-3}(t) - 3u_{-3}(t-T) + 3u_{-3}(t-2T) - u_{-3}(t-3T)]$$

(second degree pulse of height $3/8T$ and symmetrical about $t = 3T/2$)

(E.8)

IA-59,040

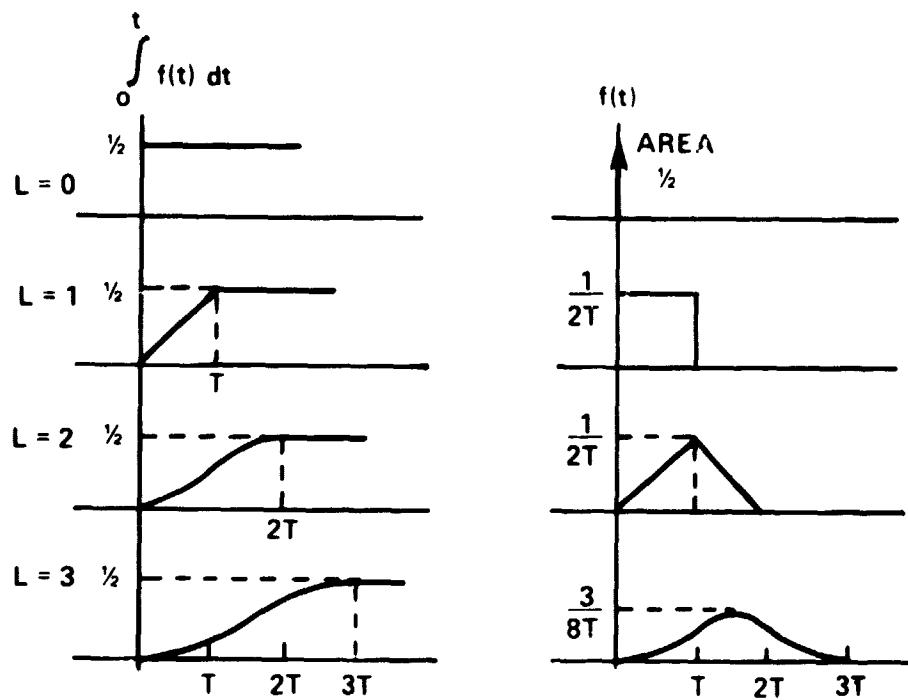


Figure E-1. Smoothing for $L=0, 1, 2, 3$; Baseband Phase and Frequency Pulses

where the unit singularity functions are given by

$$u_0(t) = \begin{cases} \infty, & t = 0 \\ 0, & \text{otherwise} \end{cases} ; \int_{-\infty}^{\infty} u_0(t) dt = 1$$

$$u_{-k}(t) = \begin{cases} \frac{t^{k-1}}{(k-1)!}, & t \geq 0 \\ 0, & \text{otherwise} \end{cases}, \quad k = 1, 2, 3, \dots . \quad (E.9)$$

In general, the LD frequency pulse can be expressed as

$$\text{LD: } f(t) = \frac{1}{2T^L} \sum_{\ell=0}^L (-1)^\ell \binom{L}{\ell} u_{-L}(t-\ell T), \quad L=0, 1, 2, \dots \quad (E.10)$$

where

$$\binom{L}{\ell} = \frac{L!}{\ell!(L-\ell)!} . \quad (E.11)$$

The proof is straightforward but somewhat involved. (This general formulation of the LD frequency pulse and the following proof is by B. E. White.)

From the definitions of the u functions of equation (E.9), it is easily verified by inspection that the $f(t)$ of equation (E.10) has $L-2$ continuous derivatives; discontinuities are possible only with step functions (or higher order singularities), and since

$$\frac{du_{-k-1}(t)}{dt} = u_{-k}(t), \quad k=0, \pm 1, \pm 2, \dots \quad (E.12)$$

step functions can arise only with the $(L-1)$ st derivative. Also, by inspection it is seen that $f(t)$ has support precisely in the interval $(0, L T)$, e.g., $f(t) = 0$ for $t > L T$ because

$$\begin{aligned}
\sum_{\ell=0}^L (-1)^\ell \binom{L}{\ell} \frac{(t-\ell T)^{L-1}}{(L-1)!} &= \frac{1}{(L-1)!} \sum_{k=0}^{L-1} (-1)^{L-1-k} t^k \sum_{\ell=0}^L (-1)^\ell \binom{L}{\ell} (\ell T)^{L-1-k} \\
&= \frac{1}{(L-1)!} \sum_{k=0}^{L-1} (-1)^{L-1-k} t^k T^{L-1-k} (-1)^L L! \begin{Bmatrix} L-1-k \\ L \end{Bmatrix} \equiv 0 \quad (\text{E.13})
\end{aligned}$$

since the Stirling number of the second kind [5]

$$\begin{Bmatrix} K \\ L \end{Bmatrix} = \begin{cases} 0, & K < L \\ 1, & K = L. \end{cases} \quad (\text{E.14})$$

Finally, the integral of equation (E.10) indeed equals 1/2:

$$\begin{aligned}
\int_0^{LT} f(t) dt &= \frac{1}{2T^L} \sum_{\ell=0}^L (-1)^\ell \binom{L}{\ell} \int_0^{LT} u_{-L}(t-\ell T) dt \\
&= \frac{1}{2T^L} \sum_{\ell=0}^L (-1)^\ell \binom{L}{\ell} \frac{1}{(L-1)!} \sum_{k=0}^{L-1} \binom{L-1}{k} (-\ell T)^{L-1-k} \int_{\ell T}^{LT} t^k dt \\
&= \frac{1}{2L!} \sum_{\ell=0}^L (-1)^\ell \binom{L}{\ell} \ell^L \sum_{j=1}^L (-1)^{L-j} \binom{L}{j} \begin{Bmatrix} L \\ \ell \end{Bmatrix}^j - 1 \\
&= \frac{1}{2L!} \sum_{\ell=0}^L (-1)^\ell \binom{L}{\ell} \ell^L \left[\left(\frac{L}{\ell} - 1 \right)^L - (-1)^L + (-1)^L \right] \\
&= \frac{1}{2L!} \sum_{\ell=0}^L (-1)^\ell \binom{L}{\ell} (L-\ell)^L \\
&= \frac{1}{2L!} \sum_{k=0}^L (-1)^{L-k} \binom{L}{k} L^k \sum_{\ell=0}^L (-1)^\ell \binom{L}{\ell} \ell^{L-k} \\
&= \frac{1}{2} \sum_{k=0}^L (-1)^k \binom{L}{k} L^k \begin{Bmatrix} L-k \\ L \end{Bmatrix} \equiv \frac{1}{2} \quad (\text{E.15})
\end{aligned}$$

from equation (E.14).

Segments of two distinct phase paths for 2D smoothing are sketched in figure E-2. The solid dots represent the underlying phases, and the vertical lines indicate the phase differences between the underlying phases of the paths. It can be seen from equations (E.2) and (E.5) that the phase difference $\Delta\phi_{ij}(t)$ between two paths is a linear function in the differences $\Delta a_n^{ij} = (a_n^i - a_{n-1}^i) - (a_n^j - a_{n-1}^j)$ between the underlying phase changes of $s_i(t)$ and $s_j(t)$ in the interval $(n-1)T < t \leq nT$, $n = 1, 2, \dots$. Note that in computing the signal space distance between the paths, cf. equation (E.5), the proper underlying phase differences Δa_n^{ij} used in calculating $\Delta\phi_{ij}(t)$ for symbol intervals shown are $0, -\pi/2, \pi/2, 3\pi/2$, and 0 , respectively, even though the absolute underlying phase difference is $5\pi/2 \equiv \pi/2$ (modulo 2π) in the last interval shown.

- LCR Frequency Pulses

An LRC frequency pulse is a raised cosine function with only one continuous derivative, extending over the L symbol interval $(0, LT)$ and having an integral of $1/2$:

$$\text{LCR: } f(t) = \frac{1}{2LT} (1 - \cos \frac{2\pi t}{LT}), \quad 0 \leq t \leq LT. \quad (\text{E.16})$$

E.3.1 R_o vs. E_b/N_o

Anderson has computed the bound parameter R_o , of equations (E.6) and (E.7), using a Markov chain model which will not be explained here. The present discussion is limited to some definitions and a sampling of relevant results.

The alphabet size M equals the number of possible underlying phase jump choices in the received signal of energy E for a given symbol interval of duration T . The modulation index is restricted to a rational number $h = r/q$, where q is the total number of underlying phases (modulo 2π), and r is an integer relatively prime to q , i.e., the greatest common divisor of r and q is unity: $\text{g.c.d.}(r, q) = 1$. This restriction on modulation index results in underlying phase changes which are multiples of $2\pi/q$ and permits a finite state Markov model when phase changes are selected at random (with the jump increment $2\pi r/q$). This also results in a phase tree which collapses into a trellis of finite states similar to those used in the Viterbi decoding of convolutional codes. [6] The parameter q is analogous to the number of quantization levels in soft-decision decoding. Since q is an indicator of receiver complexity, it should be selected to be as small as possible without compromising R_o , i.e., q should be just large enough to yield the

IA-59039

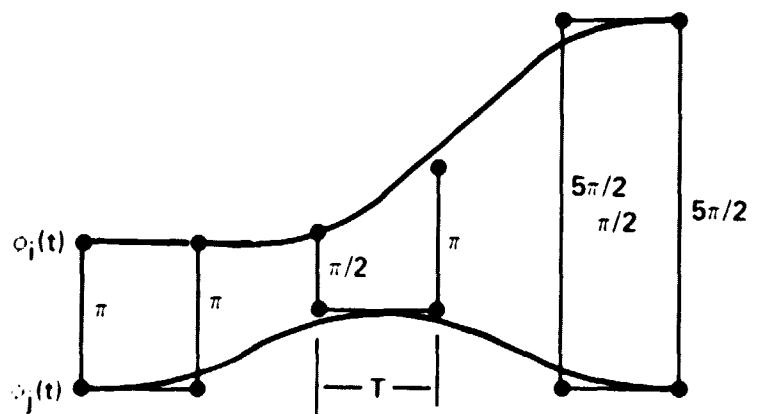


Figure E-2. Pair of 2D Smoothed Phase Trajectories Showing Underlying Phase Changes and Differences

desired performance as measured by R_o as a function of E/N_o , where N_o is the single-sided noise power spectral density.

Significant improvement in R_o is attained by increasing M from 2 to 4. As indicated in figure E-3, the dashed $M = q = 2$ curve, corresponding to antipodal signaling like binary phase shift keying (BPSK) (0D smoothing) and MSK (1D smoothing), achieves a maximum of only 1 b/symbol at about 6 dB $E/N_o = E_b/N_o$, where

$$E_b = \frac{E}{\log_2 M} \text{ (joules).} \quad (E.17)$$

Furthermore, this performance degrades considerably as the number of possible underlying phases increases. A substantial increase in R_o is evident with $M = q = 4$ and $r = 1$, even for the 2D and 3RC smoothed baseband shapes. The classical channel capacity limit

$$C = \frac{\log_2 M}{T} = \frac{1}{T_b} \text{ (b/s)} \quad (E.18)$$

of 2 b/symbol is approached fairly closely in the 10 to 12 dB range of E/N_o (or 7 to 9 dB in E_b/N_o), depending on the degree of smoothing. Recall that the 2D and 3RC frequency pulses have zero and one continuous derivative, respectively. Performance can be improved further by doubling the alphabet size to $M = 8$ where a capacity of 3 b/symbol is approached at slightly smaller E/N_o 's (E_b/N_o 's in the 5 to 7 dB range, approximately). Although relatively little decrease in R_o is observed with increased smoothing, much greater bandwidth efficiencies can result.

E.3.2 Bandwidth Efficiency

A single-sided X -dB signal bandwidth B is defined by

$$X = -10 \log_{10} \left(1 - \frac{\int_0^B S(f) df}{\int_0^\infty S(f) df} \right) \quad (E.19)$$

where $S(f)$ is the average signal power spectral density, i.e., the fractional out-of-band signal power is $10^{-X/10}$. Shannon's capacity limit formula is given by

$$C = B \log_2 \left(1 + \frac{E_b/T_b}{N_o B} \right) \quad (E.20a)$$

IA-59.041

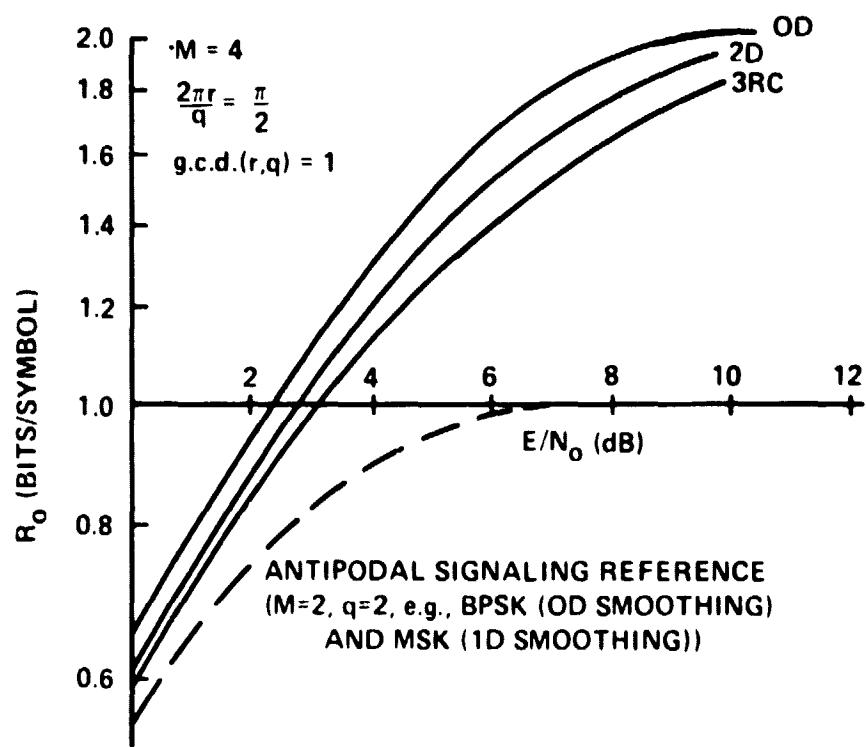


Figure E-3. Bound Parameter R_o vs. E/N_o for $M=4$ Jump Choices and Modulation Index $r/q = \frac{1}{4}$ (Jump Increment $2\pi r/q = \pi/2$) Showing Effect of OD, 2D, and 3RC Smoothing

which can be rewritten as

$$\beta = \log_2(1 + \beta E_b/N_0) \quad (E.20b)$$

defining $\beta = C/B$ and identifying the bit time T_b with $1/C$ according to equation (E.18). Plotting $2\beta = 2BT_b$ on the vertical axis to be consistent with Anderson,[4] this capacity reference is indicated in figure E-4. No modulation/coding scheme can fall below this limit.

It is useful to compare the bandwidth efficiencies of phase coded waveforms with specific baseband modulation shapes to the Shannon limit of equation (E.20). Here bandwidth efficiency is measured by the difference between the $2BT_b$ value of a particular modulation, using a specified X-dB bandwidth, and the capacity reference, both evaluated at the same E_b/N_0 value. Note that a mapping between E_b/N_0 and R_o exists through curves like those of figure E-3 for the same modulation scheme, when the bound parameter can be computed. The power efficiency is measured by the value of E_b/N_0 required. Naturally, lower values of E_b/N_0 and $2BT_b$ are desired. The bandwidth and power efficiencies for 1D and 3RC smoothing for $M = 4$ and $X = 20$ dB and 40 dB bandwidths are shown in figure E-4 for various modulation indices. It has been observed that bandwidth efficiency is improved more efficiently by decreasing h rather than increasing M above 4. As a reminder, the 1D frequency pulse is discontinuous while the 3RC frequency pulse has one continuous derivative. This accounts for the greatly improved bandwidth efficiency of the smoother pulse for $X = 40$ dB; the spectral sidelobes roll-off as 12 and 24 dB per octave for 1D and 3RC, respectively.

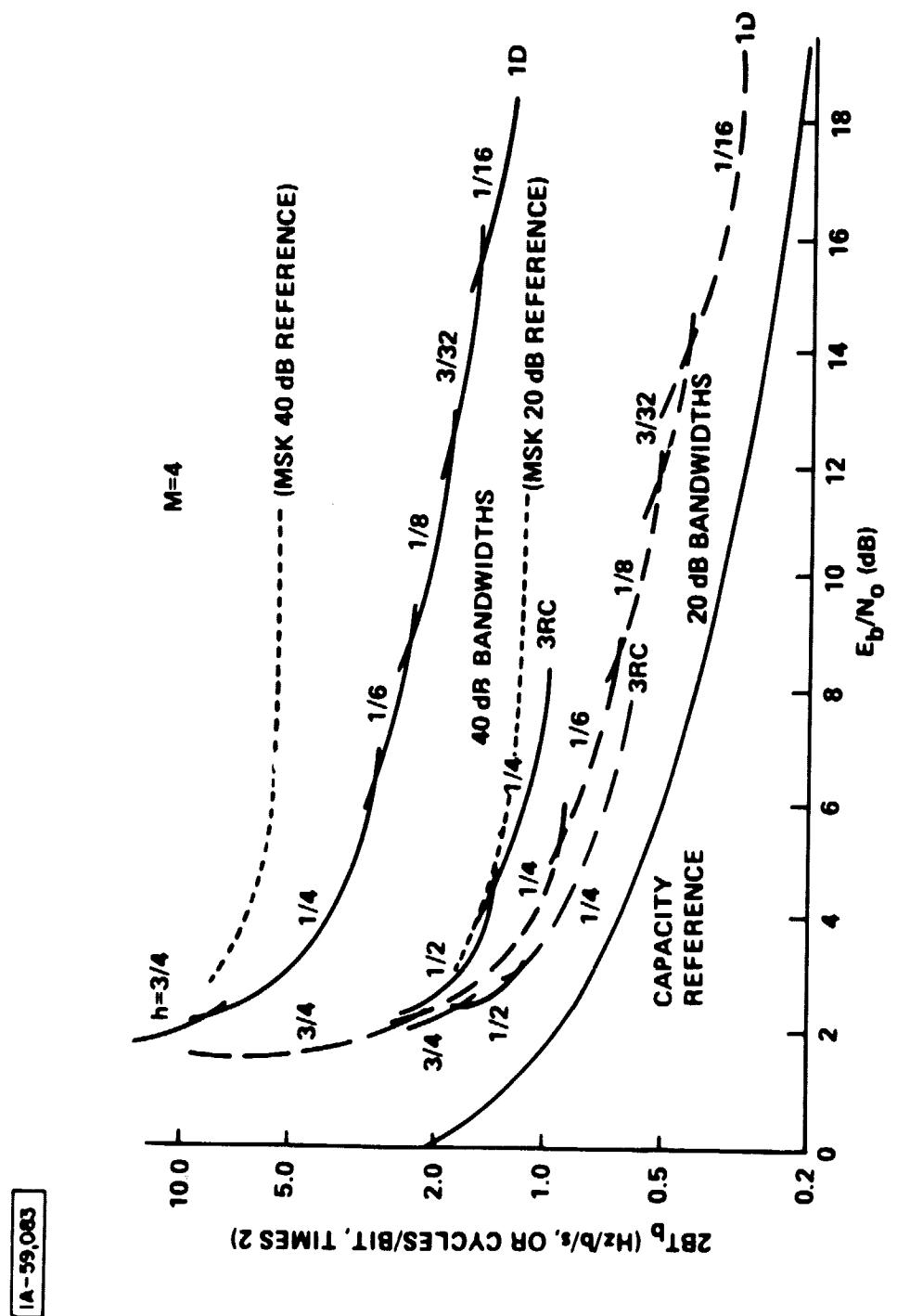


Figure E-4. Bandwidth and Power Efficiency for 1D and 3RC Smoothing, $M=4$ Jump Choices, and 20 and 40 dB Bandwidths

REFERENCES

1. Joseph L. Katz, et al., "Final Report - Application of Advanced On-Board Processing Concepts to Future Satellite Communications Systems," MTR-3787, F19628-79-C-0001, Vol. I, Bedford, MA 01730: The MITRE Corporation, June 1979.
2. William T. Brandon and Brian E. White, "Executive Summary First Year Report: On-Board Processing Concepts for Future Satellite Communications Systems," MTR-3943, F19628-80-C-0001, Bedford, MA 01730: The MITRE Corporation, May 1980.
3. Mark C. Austin and Ming U. Chang, "Quadrature Overlapped Raised-Cosine Modulation," Intern. Conf. Commun., Seattle, WA: IEEE, 8-12 June 1980, 26.7.1-26.7.5.
4. John B. Anderson, et al., "Power-Bandwidth Performance of Smoothed Phase Modulation Codes," CRL-71, Hamilton, Ontario, Canada: Communications Research Laboratory, McMaster University, March 1980.
5. D. E. Knuth, The Art of Computer Programming, Volume I, Fundamental Algorithms, Reading, MA: Addison-Wesley, 1968, 65-67, 2nd printing, 1969.
6. Stephen G. Wilson and Chung-D. Hsu, "Joint MAP Data/Phase Sequence Estimation for Trellis Phase Codes," Intern. Conf. Commun., Seattle, WA: IEEE, 8-12 June 1980, 26.1.1-26.1.5.

BIBLIOGRAPHY

Allan, R., "VLSI: Scoping its Future," *IEEE Spectrum*, Apr. 1979, p.30.

Anderson, D. B. "Integrated Optical Spectrum Analyzer: An Immanent 'Chip'," *IEEE Spectrum*, Dec. 1978, p.22

Bernhard, R., "Technology '80: Solid State Looks to VLSI," *IEEE Spectrum*, Jan. 1980, p.44.

"Big Improvement in I L Speed, Claim IBM Researchers," *Electronic Engineering Times*, Jan. 7, 1980, p.19.

Bosch, B. G., "Gigabit Electronics -- A Review," *Proc. IEEE* 67, Mar. 1979, p.340.

"Bragg Cells Advancing Rapidly, Yet Systems Application Remains Unclear," *Microwave Systems News*, Feb 1979, p.43.

Broom et al., "Model for a 15 n sec 16K RAM with Josephson Junctions," *IEEE Journal of Solid State Circuits*, Vol. SC-14, No. 4, Aug 1979, p.690.

Butler, M. B. N., "Surface Acoustic Wave Devices," *Electronic Engineering*, June 1980, p.37.

Bursky, D., "1980 Technology Forecast: 1980 ICs Will Tighten Lines to Stretch Performance," *Electronics Design*, Jan 4, 1980, p.66.

Bursky, D., "Memories and P ICs Pack More, Consume Less," *Electronic Design*, June 7, 1980, p.75.

Capece, R. P., "Semiconductors," *Electronics*, October 25, 1979 p. 104.

Capece, R. P., "Special Report: New LSI Processes," *d Electronics*, Sept. 13, 1979, p.109.

Capece, R. P., "Memories," *Electronics*, October 25, 1979 p.124.

Chang, H., "Major Activity in Magnetic Bubble Technology," *Computer Design*, Nov. 1979, p.117.

Chester, M., "Magnetic Bubble Memory Update," *Computer Design*, May 1980, p.232.

Collins, J. H., "Acousto-Optics Advance ESM," *Microwave Systems News*, Dec. 1979, p.37.

"CMOS/SOS Advances Push Signal Processing LSIs into Real Time," *Electronic Design*, Sept. 13, 1979, p.55

Connolly, R., "VHSIC Finally Gets Untracked," *Electronics*, Jan 3. 1980, p.81.

Eden, R. C. et al., "The Prospects for Ultrahigh-Speed VLSI GaAs Digital Logic," *IEEE Journal of Solid-State Circuits*, Vol. SC-14, No. 2, Apr. 1979, p.221.

"Electrically Erasable PROMs: Floating Gate Challenges MNOS," *Electronic Engineering Times*, Mar. 3, 1980, p.18.

Gossen, R. N. Jr., "100,000+ Gates on a Chip: Mastering the Minutia," *IEEE Spectru*, Mar. 1979, p.42.

Hnatek, E. R., "Semiconductor Memory Update -- Part 1: ROMs," *Computer Design*, Dec. 1979, p.67.

Hnatek, E. R., "Semiconductor Memory Update -- Part 2: RAMs," *Computer Design*, Jan. 1980, p.119.

Hnatek, E. R., "Semiconductor Memory Update -- Part 3: Higher Density Technologies," *Computer Design*, Feb 1980, p.147

Jecmen, R. et al., "HMOS II Static RAMs Overtake Bipolar Competition," *Electronics*, Sept. 13, 1979, p.124.

Keyes, R. W., "The Evolution of Digital Electronics Towards VLSI," *IEEE Journal of Solid-State Circuits*. Vol. SC-14, No.2, Apr. 1979, p.193.

Lerner, E., "Superconducting Super-Computers are on the Way," *High Technology*, Mar. 1980, p.67.

Matisoo, J., "Overview of Jospheson Technology Logic and Memory IBM J. Res. Develop. 24, Mar. 1980, p. 113.

Mavor, J., "CCD Applications Maturing," *Microwave Systems News*, Dec. 1979, p.27.

May, T. C. "Soft Errors in VLSI: Present and Future," *IEEE Transactions on Components, Hybrids and Manufacturing Technology*, Vol CHMT-2, No. 4, December 1979, p.377.

Mhatre, G., "IC Memories: RAMs," *Electronic Engineering Times*, Feb. 4, 1980, p.41.

Mhatre, G., "Memories for Microprocessors; ROMs, EPROMs, PROMs and EAROMs," *Electronic Engineering Times*, Sept. 3, 1979, p.47.

Miklosz, J. C. "Gallium Arsenide Gaining Momentum for Very Fast Circuitry," *High Technology*, Mar. 1980, p.25.

Milne, A. D., "A Survey of Present and Future Device Technologies," in "Advanced Techniques for P systems," F. K. Hanna (Ed.), Peter Peregrinus, Ltd., Stevenage, UK, 1980.

Milstein, L. B. and Das, P. U., "Surface Acoustic Wave Devices," *IEEE Communications Magazine*, Sept. 1979, p.25.

Moore, G., "VLSI: Some Fundamental Challenges," *IEEE Spectrum*, Apr. 1979, p.30.

"New Directions in Integrated Optics," *Laser Focus*, May 1980, p.40.

Posa, J. G. "Microprocessors and Microcomputers," *Electronics*, Oct. 25, 1979, p.144.

Sumney, L. W., "VHSIC: VLSI with a Vengeance," *IEEE Spectrum*, Apr. 1980, p.24.

"TELDs: Waiting in the Wings," *Microwave Systems News*, Sept. 1979, p.27.

Tsantes, "Programmable-Memory Choices Expand Design Options," *EDN*, Jan. 5, 1980, p.81.

"The Eighties: A Decade of Challenge," *Electronic Engineering Times*, Jan 7, 1980, p.1.

Wollensen, D. L., "CMOS LSI: Comparing Second-Generation Approaches," *Electronics*, Sept. 13, 1979, p.116.

Wollensen, D. L., "CMOS LSI -- The Computer Component Process of the 80's," *Computer*, Feb. 1980, p.59.

GLOSSARY

A/D	analog to digital
ACK	acknowledgment
ARQ	automatic repeat request
AWGN	additive white Gaussian noise
BFL	buffered FET logic
BM	bulk memory
BMC	bulk memory controller
BPSK	binary phase shift keying
CCD	charge coupled device
CE	constant envelope
CMOS	complementary MOS
CODEC	coder/decoder
CONUS	continental United States
CORDIC	coordinate rotations digital computer
CPS	customer premises service
CPSM	continuous phase shift modulated
CPT	customer premises terminals
CPU	central processor unit
D/A	digital to analog
DAMA	demand assignment multiple access
DCFL	direct-coupled FET logic
DFET	depletion-mode FET

DFT	discrete Fourier transform
DMA	direct memory access
DMESFET	depletion-mode MESFET
DoD	Department of Defense
EAROM	electrically alterable ROM
ECC	error-correcting coding
ECL	emitter-coupled logic
EEROM	electrically erasable ROM
EIRP	effective isotropic radiated power
EMC	electromagnetic compatibility
EMESFET	enhancement-mode MESFET
ENFET	enhancement-mode FET
EPROM	erasable PROM
FAMOS	floating-gate avalanche MOS
FDM	frequency division multiplex
FDMA	frequency division multiple access
FEC	forward error correction
FET	field effect transistor
FETLEP	link establishment protocol
FFC	FIFO controller
FFT	fast Fourier transform
FIFO	first-in, first-out (register)
FLOTOX	floating-gate tunnel oxide
FM	frequency modulation

FWHM	full width at half maximum
HJFET	heterojunction FET
IC	integrated circuit
IDT	interdigitated transducer
I L	integrated injection logic
I/O	input/output
IF	intermediate frequency
IGFET	insulated gate FET
IOSA	integrated-optic spectrum analyzer
IR	infrared
JFET	junction (i.e., PN)
LSI	large scale integration
MAC	multipoint address code
MBM	magnetic bubble memory
MDAC	multiplying digital to analog converter
MESFET	metal semiconductor (i.e., Shottky Barrier) FET
MISFET	metal-insulator-semiconductor FET
MNOS	metal-nitride-oxide-semiconductor
MODEM	modulator/demodulator
MOS	metal-oxide-semiconductor
MOSFET	metal-oxide-semiconductor FET
MPM	multi-protocol network
MSAP	mini-slotted alternating priorities
MSI	medium scale integration

MSK	minimum shift keying
MSN	magnetostatic wave
MTBF	mean time between failure
NACK	negative acknowledgment
NASA/LeRC	National Aeronautics and Space Administration/ Lewis Research Center
NMOS	n-type MOS
ONR	Office of Naval Research
PMOS	p-type MOS
PROM	programmable ROM
PRP	Packet Routing Protocol
QPSK	quadriphase shift keying
RAC	reflective array compressor
RAM	random access memory
RF	radio frequency
RMS	root mean square
ROM	read only memory
SAW(D)	surface acoustic wave (device)
SCPC	single channel per carrier
SDFL	Schottky diode FET logic
SOS	silicon on sapphire
SQPSK	staggered quadriphase shift keying
SSI	small scale integration
SYSCON	system control
TDM	time division multiplexed

TDMA	time division multiple access
TELD	transferred electron logic device
TFM	tamed frequency modulation
TTL	transistor-transistor logic
TX	(TI, T2, T3, T4) digital telephone trunk (1.544, 6.312, 44.736, 276.176 Mb/s) data rate
UHF	ultra high frequency
UV	ultraviolet
VHF	very high frequency
VHSIC	very high speed integrated circuit
VLSI	very large scale integration